

STIC-Biotech/ChemLib

From: Rao, Manjunath N.
Sent: Thursday, June 28, 2001 7:49 AM
To: STIC-Biotech/ChemLib
Subject: sequence search request for 09/668,788

From: Manjunath N. Rao
Art Unit 1652, Room 10D04
Phone: 306-5681

Date: 6-28-01

Please search the following as soon as possible for application with serial number **09/668,788**

SEQ ID NO: 1 and 3 against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO: 2 and 4 against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao
Art Unit 1652
Room CM1, 10D04
Phone 306-5681

MANJUNATH N. RAO, Ph. D
Patent Examiner
AU 1652, CM1 10D04
703-306-5681

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 1E01 TEL: 308-3534

Point of Contact:
Toby Port
Technical Info. Specialist
CMI 1E01 TEL: 308-3234

9

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Searcher: 10641 NT
Searcher Phone #: 308 3534
Searcher Location: _____
Date Searcher Picked Up: 6/29
Date Completed: 7/2
Searcher Prep & Review Time: 10
Clerical Prep Time: _____
Online Time: 10

NA Sequence (#) 2
AA Sequence (#) 2
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

STN _____

Dialog _____

Questel/Orbit _____

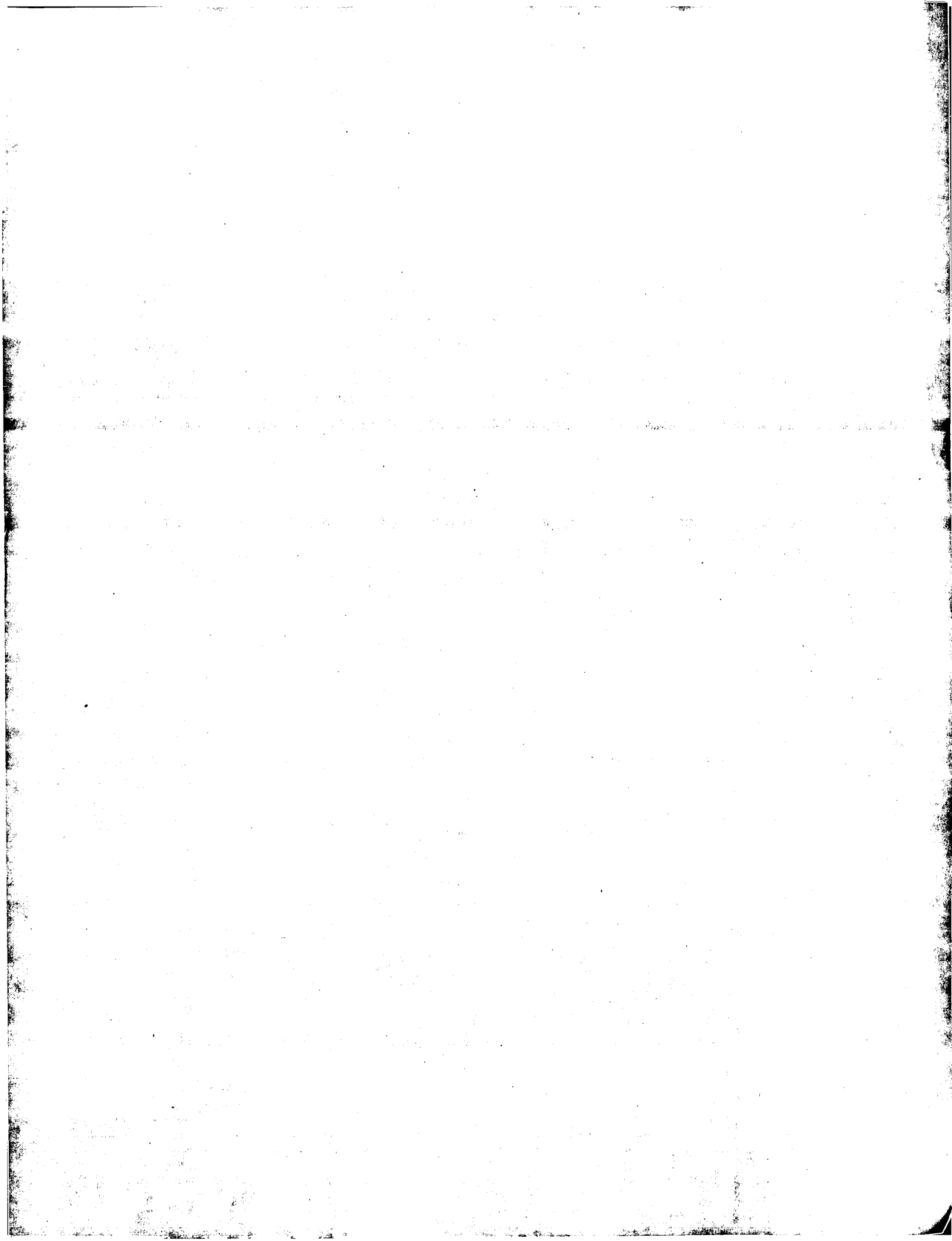
Dr. Link _____

Lexis/Nexis _____

Sequence Systems *g* _____

WWW/Internet _____

Other (specify) _____



GenCore Version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 09:33:07 ; Search time 2892.01 Seconds
(without alignments)
6145.360 Million cell updates/sec

Title: US-09-668-788-1
Perfect score: 1149
Sequence: 1 ttgataaccataaaagagt.....aagccaaagtgtatcgtaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_bai:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_bai:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_inv3:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl1:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rod:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	1149	100.0	1149	9	AX016296	AX016296 Sequence
C	2	1149	100.0	35040	2	L77246 Bacillus su
	3	1149	100.0	213680	2	Z99115 Bacillus su
4	135.4	11.8	3118	2	BSUB0012	AF270166 Staphyloc
5	135.4	11.8	3493	2	AF270394	AF270394 Staphyloc
6	120.4	10.5	7791	3	SAY14370	Y14370 Staphylococ
7	118	10.3	975	9	AX016297	AX016297 Sequence
8	69.6	6.1	292550	2	AP001513	AP001513 Bacillus

9 53.8 4.7 13408 1 AE001958
c 10 49 4.3 7218 10 I66494
11 48 4.2 1266 10 E14811
12 48 4.2 1266 10 E14811
c 13 44.4 3.9 150583 88 AC026884
c 14 44.4 3.9 170934 79 AL354684
c 15 44.2 3.8 55593 2 BSJ2571
c 16 44.2 3.8 208230 2 BSUB0008
c 17 44.2 3.8 216750 2 BSUB0007
c 18 43.2 3.8 1141 10 AX083744
c 19 43.2 3.8 9070 13 AF148542
c 20 43 3.7 1569 15 SOL249607
c 21 42.8 3.7 2055 12 AB047476
c 22 42.6 3.7 158293 85 AC004746
c 23 42.6 3.7 172209 75 AC074253
c 24 42.2 3.7 5742 96 PYEL6
c 25 42 3.7 137641 78 AF273085
c 26 41.8 3.6 104992 60 AC005504
c 27 41.8 3.6 150467 87 AC009541
c 28 41.8 3.6 169546 60 AC004157
c 29 41.8 3.6 172237 88 AC083869
c 30 41.6 3.6 187610 67 AC022218
c 31 41.4 3.6 5220 2 BBHGBABCD
c 32 41.4 3.6 30303 6 CELC18B2
c 33 41.4 3.6 52971 1 AE001584
c 34 41.4 3.6 157848 65 AC018455
c 35 41.4 3.6 167032 87 AC012157
c 36 41.4 3.6 171255 86 AC007912
c 37 41.2 3.6 138890 91 CANS01RGD
c 38 41.2 3.6 160000 92 AC037443
c 39 41.2 3.6 151961 71 AC016198
c 40 41.2 3.6 161474 64 AC016198
c 41 41 3.6 49898 70 AC027136
c 42 41 3.6 127447 86 AC006344
c 43 41 3.6 190912 90 AL357559
c 44 41 3.6 198479 73 AC068483
c 45 40.8 3.6 140908 74 AC069516

ALIGNMENTS

RESULT 1
AX016296 1149 bp DNA PAT 07-SEP-2000
LOCUS Sequence 1 from Patent WO9949052.
DEFINITION AX016296
ACCESSION AX016296
VERSION AX016296.1 GI:10041859
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
Bacillus (bases 1 to 1149)
Zaehringer, U., Heinz, E., Jorasz, P. and Wolter, F.P.
TITLE Processive glycosyltransferase
PATENT: WO 9949052-A 1 30-SEP-1999;
JOURNAL FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
(DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
UND VERWERB (DE)
FEATURES
source
Location/Qualifiers
1. .1149
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 380 a 217 c 244 g 308 t
ORIGIN
Query Match 100.0%; Score 1149; DB 9; Length 1149;
Best Local Similarity 100.0%; Pred. NO. 3.3e-265;
Matches 1149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ttgaataaccataaaagagtattattttgactgacaaattacggaaatggacatgtgcag 60

Db 1 TTGAATACCAATAAAAGAGTATTATTTTGAAGTCAAAATACGGAATGGACATGTGCAG 60
Qy 61 gtacgcaaaacactttatgaacaatgttacggctcgcttttcagcatgaacagttctt 120
Db 61 GTAGCCAAAACACTTTATGAACAATGTGACGGCTCGGCTTTTCAGCATGTACAGTTCT 120
Qy 121 aattgtaccagaagtcaaatccgattgtttcagagtgtaactcaatcacctttatttaa 180
Db 121 AATTGTACCAAGAGTCAAAATCCGATTGTTTCAGAGSTAACCAATACCTTTATTTAAA 180
Qy 181 agttctcaatcgggaaacagttttttatcgtttttttattacggagttgacaaaatctat 240
Db 181 AGCTTCTCAATCGGAAACAGTTTATCGTTTGTGTTTATTCAGGAGTTGACAAAATCTAT 240
Qy 241 aataaagctaaattcaatatttactttaaaatgggttaataaaagattggcggaactgtc 300
Db 241 AATAAGCTAAATCAATATTTACTTTAAATGGGTAAATAAAGATGGCGGAACTGTGC 300
Qy 301 gatgaacatcagcccgatattattattacattccgatattccgatcgtgcgcggaatac 360
Db 301 GATGAACATCAGCCGATATTATTATTAATACATTTCCGATGATCGCTGCGGGAATAC 360
Qy 361 agacgcgaactggaagagtcattcctcactcaactcaagttatgactattttttcttcat 420
Db 361 AGACGCCGAATCGGAAGAGTCATTCTTACCTTCAACGTTATGACTGATTTTGTCTTCAT 420
Qy 421 aaatttgggttcaacgaaacgtgataaattattattgtgcgcagacagattacgtgaaggaa 480
Db 421 AAATTTGGGTTCACGAAACGTTGATATAATATTATGTGGCAGAGATTACGTGAAGGAA 480
Qy 481 aaactgtgagatcgccactcattcaagaacattgataaaatattattgtgcgcagacagattacgtgaaggaa 540
Db 481 AAATGTGGAGATCGGCACCTCATCCAAAGCAATGTAATAATACAGGAATTTCCAAATCAGG 540
Qy 541 ccgaatttaagaatccatgcctgttggccgatataataaaagtcacacttttccacca 600
Db 541 CCGCAATTTGAAGAATCCATGCCCTGTTGGCCCGCATATATAAAGATACATCTTTCCACCA 600
Qy 601 aacaaaaagtgtcttctgatcatgagcaggtgtcacggtgtattataaagaacgttaaaagag 660
Db 601 AACAAAAAGTGTCTTGATCATGTCAGAGTGTCTCACGGTGTATTAAAAAGACGTAAAAAGAG 660
Qy 661 ctgtgcgaacacctgtgacagatgacaaagtgaagttagtctgctgctgcgggaaaaat 720
Db 661 CTGTGCGAAAACCTGTGCAAGGATGACCAAGTAGTGTGCTGCTGCGGGAATAAT 720
Qy 721 acgcttttaagaatctttgagtcgcttgaagcgaataatggtgacaaatttaaaagt 780
Db 721 ACGGCTTTAAAGAAATCTTTGAGTCCGCTTGAACGGGAAATGCTGACAAATTTAAAGTT 780
Qy 781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgtaccagaag 840
Db 781 CTGGGCTATGTGGAGCGCATTTGATGAGCTATTTCCGGATTCACAGATTGATGATTACCAAG 840
Qy 841 cccggcgccattactttgaagaagccacacccattgagtcctgctcattctgtacaaa 900
Db 841 CCGGCGGCATTACTTTTACAGAGCCACGCCCATTTGAGTGCCCTGCTGCTGTACAAA 900
Qy 901 cccgtgcctggcaggaagaaataatgcaactcttttgaagccgagcgtgcgcattc 960
Db 901 CCGGCTGCTGCGCAGGAAAAAAGAAATGCAAACTCTTTTGAAGACCGCGGAGCTGCCATC 960
Qy 961 gtttgtaaccgtcatgaagagattctcgagtcagtcacttccctcttgcagatgaagat 1020
Db 961 GTTGTGAACCCGTCATGAAGAGATTCTCGAGTCAGTCACCTCCCTTCTTTCAGATGAAGAT 1020
Qy 1021 acctgtcatcgcataagaagaaacatttaagacattttagcaaaactcctctgaagt 1080
Db 1021 ACCTTGCATCGCATGAAGAAACATTAAAGACCTTCAATTATAGCAAACTCCTCTGAAGTG 1080
Qy 1081 atttagaggatatcctgaaggaatcagaatgatgacgcgcaacaaagcgaagc 1140

Db 13796 CTATCGTAA 13788

RESULT 3

BSUB0012

LOCUS

DEFINITION

BSUB0012 213680 bp DNA BCT 26-NOV-1997

VERSION

299115

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

ABSTRACT

NOTES

COMMENTS

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

/translation="MDSKKHKIINGYHKNCISCKSWLPATEENFYSVKKNKDLGHSY
 CKACVLKAKESMLKNYDQLEMRERENLLPGKEAKKKYNSSLSKKKTKQYIQEKNK
 LKLNTRQDAHKHNTDQWCKDYFNKCSYGLKIEDHKILFKGTIYQSDPH
 KEVHDHKGANDISNCIPACKSSKSHDFAEEMYNSSSNKFNFSERLLKTKLEWLNRFK
 EERQDSEWRTKG
 complement(822. .1025)
 /gene="yocq"
 complement(822. .1025)
 /gene="yocq"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13978.1"
 /db_xref="GI:2634480"
 /db_xref="SPTREMBL:O34326"
 /translation="MKLIERTYERLYIYSDDEARKHFVSELKPKGMVNDNCWGI
 DLNQIKQFYRKRELTDDYVFN"
 complement(1045. .1560)
 /gene="yocq"
 complement(1045. .1560)
 /gene="yocq"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13979.1"
 /db_xref="GI:2634481"
 /db_xref="SPTREMBL:O34359"
 /translation="MKNPTMLKDKLLAVIELITKENKYRFTGALGTDOACWC
 VHLLKAKHPIKNIATPFREKQKVSADQRMKMLNVADEIINVEEDKYKVSQE
 KPELDSFAPKQKREYMIHSEIAVAVYDGSKTRNCLNAYAKTKYLGHQLRLHPDF
 NFELDTYFVG"
 complement(1771. .1965)
 /gene="yocq"
 complement(1771. .1965)
 /gene="yocq"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13980.1"
 /db_xref="GI:2634482"
 /db_xref="SPTREMBL:O34339"
 /translation="MKSIILPEENTKITVCKPINDESNTKVIAIYDREPEAFWV
 HLSGNDLFDVNDHEVIVEYE"
 complement(2014. .2466)
 /gene="yocq"
 complement(2014. .2466)
 /gene="yocq"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13981.1"
 /db_xref="GI:2634483"
 /db_xref="SPTREMBL:O34999"
 /translation="MKRFILVLSFLSIIVATQTNASPMPCSVILEPVDKNLKNAG
 VALIKVQLNPPSAARTNISILAVHLPAPSGYNDYEGFATKPGGISWRFKLYPTP
 EEPSPWAGRIDTISAEMKNVQVRLSNSTQKLGPSILTKNIESCY"
 complement(2549. .2806)
 /gene="yocq"
 complement(2549. .2806)
 /gene="yocq"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13982.1"
 /db_xref="GI:2634484"
 /db_xref="SPTREMBL:O35030"
 /translation="MERKLLKEITFDLKNKIKPGTKVQVTKYKTEMYTHNGEVII
 NNGEKFYNYVDKEGYVGHCHVNALDKNYPDSLIVEIKSK"
 complement(2851. .3054)
 /gene="yocq"
 complement(2851. .3054)
 /gene="yocq"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13977.1"
 /db_xref="GI:2634479"
 /db_xref="SPTREMBL:O34581"

```
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB13983.1"
/db_xref="GI:2634485"
/db_xref="SPTREMBL:O34584"
/translation="MKFAPMDKVFYKTAASHLNKRLTLKKRVPELDDPLLGECEWFEED
GLRQFDWEEYEFVAPRPFNWD"
complement(3063..3227)
/gene="yoqB"
CDS
complement(3063..3227)
/gene="yoqB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB13984.1"
/db_xref="GI:2634486"
/db_xref="SPTREMBL:O34429"
/translation="MENVLYGVYSTRSRKFCFGLIEPSKTKARKELFNRICTDAYKW
RFEIRKIRK"
complement(3281..4036)
/gene="yoqD"
CDS
complement(3281..4036)
/gene="yoqD"
/function="unknown"
/note="similar to phage-related DNA-binding protein
anti-repressor"
/codon_start=1
/transl_table=11
/protein_id="CAB13985.1"
/db_xref="GI:2634487"
/db_xref="SPTREMBL:O34449"
/translation="MESVLTVEQQLLVDSREVAVMGKRHTDLRSIDGYVAILL
NAKLRSEFLESTYKDATGRSLKHFLTRKCDVMANKMGAKVLTAOVSKFEE
MEKALKARSLIDITLDNEDERATAYTERKAKRELQELTLAEPKVYKDYRLNTD
GLMKIGQVAKATIGKMGONNLFRELRENKVLIDTNKNAPQKIVERGFFQVKTOET
SVGIKTIIVTPKGADFIYDLKHKHKEIAS"
complement(4077..4484)
/gene="yoqC"
CDS
complement(4077..4484)
/gene="yoqC"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB13986.1"
/db_xref="GI:2634488"
/db_xref="SPTREMBL:O34834"
/translation="MEVNVKTNQREKFIIRNGIPYDELDTQMLHILDLNFKLGLKTRH
CCFGRHPYEIQWFEVNEKEDQILELAELAGREWKGLQLSFSKWAEFSPLMFNWS
LVLSKRFNPEDPNKYRLRSVEEFESYAARK"
complement(4491..4829)
/gene="yoqB"
CDS
complement(4491..4829)
/gene="yoqB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB13987.1"
/db_xref="GI:2634489"
/db_xref="GI:2634489"
```

```
Query Match 100.0%; Score 1149; DB 2; Length 213680;
Best Local Similarity 100.0%; Pred. NO. 5.7e-265;
Matches 1149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgaatacaataaagaagtattattttgactgcaataattacggaatggacatgtgcag 60
Db 110216 TTGAATACATAAAGAAGATTAATTTTGACTCAAATTACGGAATGGACATGTGCAG 110275

QY 61 gtacgcaaaacactttatgaacaatgtgtacggtctcggtttcagcatgtaacaattct 120
Db 110276 GTAGCCAAAACACTTTATGAACAATGTGTACGGCTTCGCGCTTTCAGCATGTAAACAGTTCT 110335

QY 121 aatttgtaccaagagtcacaaatccgattgtttcagagggtacactcaatacctttattaaaa 180
```

```
Db 110336 AATTTGTACCAAGAGTCAAATCCGATTGTTTTCAGAGTAACCTCAATACCTTTATTTAAAA 110395
QY 181 agcttctcaatcgggaaacagttttatcgttttttattacgaggttaacaaaatctat 240
Db 110396 AGCTTCTCAATCGGAAACAGTTTTATCGTTTGTATTTATACGGAGTTGACAAAATCTAT 110455
QY 241 aataaacgttaaatcaatattacttttaaatgggtaataaaaagattggcggaactgtc 300
Db 110456 AATAAACGTAAATCAATATTACTTTAAATGGTAATAAAAGATTGGCGAATGTGC 110515
QY 301 gatgaacatcagcccgatattattattaatacatttcgatgatcgtcgtgcgggaatac 360
Db 110516 GATGAACATCAGCCGATATTATTATTATACATTTCCGATGATCGTCGCGCGAATAC 110575
QY 361 agacgcgaaactggaagagtcattccctcaactcaactcaacttattgactgattttgtcttc 420
Db 110576 AGACGCGAACTGGAAGAGTCATTCACCTTCAACGTTATGACTGATTTTGTCTTCAT 110635
QY 421 aaaatttgggttcacgaaacgttgataaataattattatggtgcagagattacgtgaagaa 480
Db 110636 AAAATTTGGTTACGAAACGTCGATAAATATTATGTGCGACAGATTACGTGAAGGAA 110695
QY 481 aaactgtggagatcggcactcatccaagcaatgtaaaaatcacagaattccaatcagg 540
Db 110696 AAACGTCTGGAGATCGGCACCTCATCCAAGCAATGTAAAAATCACGGAATTCCAATCAGG 110755
QY 541 ccgcaatttgaagaatccatgcctgttgcccgatataataaaagtagacaattcttcacca 600
Db 110756 CCGCAATTTGAAGAATCCATGCTTGTGGCCGATATATAAAAAAGTACAAATCTTTCACCA 110815
QY 601 aacaaaagtgctctgatcatggcaggtgctcacggtgtattaaagaacgttaaaagag 660
Db 110816 AACAAAAAGTCTCTCATCATGCGAGTGCTCACGGTGTATTAAAGAACGTAAAGAG 110875
QY 661 ctgtgcgaaaacctgtcaagatgaccaaagtgcgaagttagtctgtgctgctgcgggaaaaat 720
Db 110876 CTGTGCGAAAACCTTGTCAAGGATGACCAAGTGAAGTAGTGTGCTGTGCGGGAAAAAT 110935
QY 721 acggttttaaaagaattctttagtgcgttgaaagcggaataatggtgacaaattaaaagtt 780
Db 110936 ACGGCTTTAAAAAGAATCTTTGAGTGCCTTGAAGCGGAAAAATGGTGCACAAATTTAAAGTT 110995
QY 781 ctggactatgtgagcgcattgatgatctatttcggatcacagattgcatgattaccacag 840
Db 110996 CTGGGCTATGTGGAGCGCATTTGATGAGCTATTTTCGGATCACAGATTGCATGATTACCAAG 111055
QY 841 cccgcggtcattactttgacagaagccacagccattggagtgccctgattctgtacaaa 900
Db 111056 CCGCGCGCATTTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCTCATTTGTACAAA 111115
QY 901 cccgtgcctggccaggaaaaaagaaatgcaaatcttttgaagaccgagcgagctgccatc 960
Db 111116 CCGCTGCTGGCCAGGAAAAAGAAATGCAAACTCTTTTGAAGACCGGGAGCTGCCATC 111175
QY 961 gttgtgaaccgtcatgaagagattctcagtcagtcacttcccttcttgcagatgaagat 1020
Db 111176 GTTGTGAACCGTCATGAAGAGATTCTCGAGTCAGTCACCTTCCTTCTTTCAGATGAAGAT 111235
QY 1021 acctgcacatgcgatgaagaaaaacataagagaccttcattagcaaacctcctcgaagtq 1080
Db 111236 ACCTTCATCGCATGAAGAAAAACATTAAGGACCTTCATTTAGCAAACTCCTCTGAAGTG 111295
QY 1081 attttagagatctcctgaagaatcagaaatgatgaccgccaacaaacaaagccaaagtg 1140
Db 111296 ATTTTAGAGATATCTCTGAAGGAATCAGAATGATGACCGCCCAACAAAAAGCCAAAGTG 111355
QY 1141 ctatcgtaa 1149
Db 111356 CTATCGTAA 111364
```

RESULT 4

AF270166	3118 bp	DNA	BCT	01-AUG-2000		
LOCUS	3118 bp	DNA	BCT	01-AUG-2000		
DEFINITION	Staphylococcus epidermidis strain SR1 clone step.1051e11 genomic sequence.					
ACCESSION	AF270166					
VERSION	AF270166					
KEYWORDS	AF270166.1	GI:9624074				
SOURCE	Staphylococcus epidermidis.					
ORGANISM	Staphylococcus epidermidis					
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus.					
AUTHORS	1 (bases 1 to 3118) Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelsen,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mammo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.					
TITLE	Transposon-mediated sequencing of the Staphylococcus epidermidis genome					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 3118)					
AUTHORS	Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelsen,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mammo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.					
TITLE	Direct Submission					
JOURNAL	Submitted (22-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA					
FEATURES	Location/Qualifiers					
source	1..3118 /organism="Staphylococcus epidermidis" /strain="SR1" /db_xref="taxon:1282" /clone="step.1051e11"					
BASE COUNT	1048 a	480 c	492 g	1098 t		
ORIGIN						
Query Match	11.8%;	Score 135.4;	DB 2;	Length 3118;		
Best Local Similarity	46.7%;	Pred. No. 4.3e-22;				
Matches 510;	Conservative	0;	Mismatches 566;	Indels 15; Gaps 2;		
QY	10	aataaaagagtataattttgactgcgaattacggaatgacatgtgcaggtagcaca 69				
Db	287	AATAAAAAAGATATTGATTACTGGATCATTTTGGTAACGGTCATATGCAAGTCACGCA 346				
QY	70	acactttatgaacaatgtgtacggctcggttcacgatgtaacagt-----ttcta 123				
Db	347	AGTATTGTCAACCAATTGGAATGAGATGAATCTCAATCATTTTACGTATTCACATGAT 406				
QY	124	ttgtaccagaagtcaaatccgattgtttcagagggttaactcaatcacctttattt 183				
Db	407	TTGTTTATGGAAGCTCATCCATTATGACCTTCTATATGTAAAGATGGTATATCAATAGC 466				
QY	184	ttctcaatcgggaaacagttttatcgttgttttatttaacgaggttgacaaaatc 243				
Db	467	TTTAAATATTTTGAAGAAATACATATAAACGATTTTACTATAGTCGCCCTAATGAGCTCGAT 526				
QY	244	aaacgtaaattcaatatttactttaaaattggtaataaaaagattggcgaaactgtcgat 303				
Db	527	AAATGTTTTTATAAATATTA-----TGGATTAAATAAATCATCATCTACTTATT 577				
QY	304	gaacatcagcccgatattattataataatccgatgatcgtcgcggaataacaga 363				
Db	578	AAGAAGAAAGCCTGATCTCATATTATTAAACATTTTCCAAACACCTGTGATGTCAGTGTGACC 637				
QY	364	cgccgaactggagagtcattccacttcaacgttatgactgactgtttttgtcttcaataa 423				
Db	638	GACAATTTAATAATAATATCCCTATTTCGACAGAGTTATGACAGATTATTCGCATGCATATAA 697				
QY	424	atttgggttcacgaacacgtgggaataattattgtgcgacagattacgtggaaggaaaaa 483				

```

Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE
Direct Submission
JOURNAL
Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
Location/Qualifiers
source
1. 3493
/organism="Staphylococcus epidermidis"
/strain="SR1"
/db_xref="taxon:1282"
/clone="step.4046a02"
BASE COUNT
1172 a 534 c 547 g 1240 t
ORIGIN

```

	Query Match	11.88;	Score 135.4;	DB 2;	Length 3493;	
	Best Local Similarity	46.7%;	Pred. No. 4.3e-22;			
	Matches 510; Conservative	0;	Mismatches 566;	Indels 15;	Gaps 2;	
Qy	10 aataaaagagttaaattgactgcgaatcacggaatggacatgtgcagtgtagcgaaca	69				
Dd						
Qy	396 AATAAAAAGATATTGATTTACTTGGTCATTTGGTAACGTGCATATGCAAGTCACGCCAA	455				
Dd						
Qy	70 acactttatgaacaatgttgtcgctcggcttcaggcatgttaaacagt-----tcttaat	123				
Dd						
Qy	456 AGTATTGTCTAACCAAAATGAATGAGATGAATCTCAATCATTTTATCAGTCATTTCAACATGAT	515				
Dd						
Qy	124 ttgtaccaagagtcgaatccgatgtttccagagtgtaactcaataacctttattaaaaagc	183				
Dd						
Qy	516 TTGTTTTATGGAAGTCATCCAATTTATGACTTCTATATGTAGAATAATGTTATCAATAGC	575				
Dd						
Qy	184 ttctcaatcgggaacagtttatogtttggttttattaccgaggttgacaaaatctcataat	243				
Dd						
Qy	576 TTTAAATATTTTAGAAAATACATATAAAGCATTTTACTATAGTCGGCCCTAATGAGTCGAT	635				
Dd						
Qy	244 aaacgtaaatcaatatatttaactttaaaatggtaaaaaagattgggcgaactgtcगत	303				
Dd						
Qy	636 AAATGTTTTTATAAATATTA-----TGATTAATAAATAACTCATCACTTACTTATTT	686				
Dd						
Qy	304 gaacatcagccgatattattattaacattccgatgatcgctgctgccggaatacacaga	363				
Dd						
Qy	687 AAAGAAAGCGCTGATCTCATATTATTAACAATTTCCAACACCTGTGATGTCAGTGTGACC	746				
Dd						
Qy	364 cgcggaactcgaagagtcattoctaccctccaagttatgactgattttgttctcataaa	423				
Dd						
Qy	747 GAACAATTTAATATAAAATATCCCTATTCCGACAGTTATGACAGATTTCGCAATGATCAA	806				
Dd						
Qy	424 attgggttcagcaaacggtgataaattatgtggtgcagagattacgtgaaggaaaaa	483				
Dd						
Qy	807 AATTGGATTACACCATATTCAAAGATATATGTAGCAACAAAGATACATAAGATGAT	866				
Dd						
Qy	484 ctgctggagatcggcactcatccagcaatgtataaaatccagggaattccactcggccg	543				
Dd						
Qy	867 TTCATTGAAGCTGGTGTCTCTTCATATATAAAGTCACGGGATTCCTATTGCTGAT	926				
Dd						
Qy	544 caatttgaagaatccatgcctgttgcccagatataaaaaagtaacaattttccacaaac	603				
Dd						
Qy	927 AAAATTTGAAGAAATCTATTGATAAAGAAAGTGGTTATGCAACACACATTTAGACCTTCA	986				
Dd						
Qy	604 aaaaaagtgcttcgatcatgcaggtgtcaggtgtattaaagaacgtataaagagctg	663				
Dd						
Qy	987 AACCTACTATATTAAATGTCACGAGTGCATTTGGTGTGTTTCAAAAGGCTTTGACATATG	1046				
Dd						
Qy	664 tgcgaaaaccttgcgaagatgaccaagtgcgaagtgtgtcgtgtgctgccccaaatcag	723				
Dd						
Qy	1047 ATTAATAATATTTTAGAAAAAAGTCCAAATTCGCAAGTGGTCATGTTGTGGAGCTAGT	1106				
Dd						
Qy	724 gctttaaaaagaatctttgagtcgcttgaagcggaaaaatgggcacaaattaaagctcg	783				
Dd						
Qy	1107 AAGGAACCTTAACGGTTTCAATAAAGCTAAGTTCAAGAGATAATCCAAAGTAAATATTA	1166				
Dd						
Qy	784 ggctatgtggagcgcattgatgagctatttcggatcacagattgatgattaccagccoc	843				
Dd						

RESULT	6
SAY14370	7791 bp DNA BCT 24-JUN-1998
LOCUS	Staphylococcus aureus RF3, mureE, ypfp genes.
DEFINITION	Y14370
ACCESSION	Y14370.1 GI:3256221
VERSION	mure gene; peptide chain release factor 3; RF3 gene;
KEYWORDS	UDP-N-acetylmuramyl-tripeptide synthetase; ypfp gene.
SOURCE	Staphylococcus aureus
ORGANISM	Staphylococcus aureus
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;
AUTHORS	Bacillus/Staphylococcus group; Staphylococcus.
TITLE	1 (bases 1 to 7791)
JOURNAL	Ludovice,A.M., Wu,S. and de Lencastre,H.
REFERENCE	Molecular cloning and DNA sequencing of the Staphylococcus aureus
AUTHORS	UDP-N-acetylmuramyl tripeptide synthetase (mure) gene, essential
TITLE	for the optimal expression of methicillin resistance
JOURNAL	Microb. Drug Res. 4, 85-90 (1998)
REFERENCE	2 (bases 1 to 7791)
AUTHORS	Ludovice,A.M.
TITLE	Direct Submission
JOURNAL	Submitted (30-JUL-1997) A.M. Ludovice, Instituto de Tecnologia
REMARK	Olimica e Biologica, Universidade Nova de Lisbon, R. da Quinta
COMMENT	Grande 6, Apartado 127, 2780 Oeiras, PORTUGAL
FEATURES	Revised by author 16-JAN-98
source	Related sequence: L77246. Location/Qualifiers 1..7791 /organism="Staphylococcus aureus" /strain="COL" /db_xref="taxon:1280" complement(1435..3114) /gene="RF3" complement(1435..3000) /gene="RF3" /codon_start=1 /transl_table=11 /product="peptide chain release factor 3" /protein_id="CAA4739.1" /db_xref="GI:3256222" /db_xref="SPTMBL:O86490" /translation="MNLQEVSRKTFALISHPDAGKTTLTKLYLFSGAIREACTV GKLVLNVRVMWVBOERGISVTSSWMQFDYDVEINILDPGGHEFSDFTYRLMA VSVAVMIDCAQGVETPLLEFKVCCKMRGITFTINKLDVRGKEPFLLEDIEETLN IETYPMNWIGMQSFGLIDRKSRTIEPRDEENILHLDDFELEDOHALTNDSDFE QAIEELMWEEAGEANDALLUSDGLTPVFPGSALANGFQVFNLAYDYDFAMPNARQ TKENVEVPDSFSFGIFPKIQANMDPKHRDIAMRVVSYGAKTYQSDDLQPQTPEI RSVHQLWQTIKLVNHAVAGDIIGLYDTGNQIGDLTVLGKGKTYSQDLPQFTPEI FMVSKSNMYKKQHFKHGTEQLVQEGAIQYKRTLHTNQIILGAVGLOQEFVEFHRMK


```
ENVDVMPVGRKIARVDIENEDQITDKMNTSRSLVKRDYDDLFLPENEFATRWE
EKKFPIKLYSL"
RBS complement(3009. .3013)
-10_signal /gene="RF3" complement(3085. .3091)
-35_signal /gene="RF3" complement(3109. .3114)
gene /gene="RF3" complement(3244. .4784)
CDS /gene="mure" complement(3244. .4722)
/codon_start=1
/transl_table=11
/product="UPP-N-acetylmuramyl-tripeptide synthetase"
/protein_id="CAA74740.1"
/db_xref="GI:3256223"
/db_xref="SPTREMBL:O86491"
/translation="MQVRLKVKVKKRVLSLEQQIDITTSRTAREGSIFVASVGY
TVDSHKFCQNAVADQGLVNVNKEQSLPANVTQVVVPDRLVASILAHPTLYDPSHOL
VTFGVGTNGKTSIATMIHLQRLKQNSAYLGTNGFOINETKTGANTPTETVSLTK
KKEADVDAESMTLSESSHLVGLRSEVDFVALFESNLTQDHLDFHGTMEAYGHAK
SLLFSQGLDLSKEKIVLVNDDSFSEYLRTPYEVFSIGIDEQAFMAKNIEQLQ
GVSFDFPFYIPVPSKPYVGFKNISNMAAMIAVWSKTSLETIIKAVENLPEVGR
LEVLDPSLDIIDIYAHFADGNKLIIDAVQFVKOKLFLVGMAGERDLTKTPEMR
VACRADYVFTPDNPANDPKMLTAEALKGATHQNYIEFDRAEGIKHAIDIAEPDGT
VVLAKGREPIYIMPGHIKVPHRDDLIGLEAYKFKGGGVPD"
RBS complement(4736. .4740)
-10_signal /gene="mure" complement(4758. .4763)
-35_signal /gene="mure" complement(4779. .4784)
gene /gene="mure" 5086. .6335
-35_signal /gene="yfp" 5086. .5092
-10_signal /gene="yfp" 5107. .5112
RBS /gene="yfp" 5145. .5150
CDS /gene="yfp" 5160. .6335
/transl_table=11
/codon_start=1
/protein_id="CAA74741.1"
/db_xref="GI:3256224"
/db_xref="SPTREMBL:O86492"
/translation="MVTONKKILLITGSEFGNGHMOVTQSIYNQNDMNLHLVIEHD
LPWEAHPILTSCKKWINSFYFRNMYKFYSRDRKDKCFYKIYGLNKLNLILIK
EXPDLILFFTPVMSVLQFNINIPVATYDRLKHNWITPYSTRYYVATKTKQ
DFIDVIGSTVKVGPIDNKFETPINOKWLDNLDPDKOTILMSAGAFVSKGF
DMITDILAKSAAQVMTCGSKSKELKRLMYLILGTYKHNENWASSQL
MITKPGIITGPFARCIPIWFLNAPGOELENAFYFEKGFKIADTPEEAIKIVAS
LTNGNQLNMLSTMBODKIKYATOTICRDLDLGHSSQPOEIIYKVPYARFVK"
BASE COUNT 2517 a 1452 c 1130 g 2668 t 24 others
ORIGIN
```

Query Match 10.5%; Score 120.4; DB 3; Length 7791;

Best Local Similarity 48.3%; Pred. No. 1.9e-18;

Matches 473; Conservative 0; Mismatches 486; Indels 21; Gaps 4;

QY 10 aataaagagatataattttgactgcaaatcaggaatggacatgtagcagtagccaaa 69

Db 5172 AATAAAGATATTGATTATTACTGGTCAATCGGTAAACGGTCATATGCAAGTTACACAG 5231

QY 70 acatttataacaatgtgtacgcgcgttcagcatgtaaacagtttct-----aat 123

Db 5232 AGTATCGTTAATCACTTAATGATATGAATCTAGACCATTTAAGCGTCATTGAGCAGCAT 5291

QY 124 ttgtaccagagcacaatccgattgttttcagaggttaactcaatccttttatttaaaagc 183

```
Db 5292 TTATTTATGGAAGTCATCCAAATTTTTCACCTTCTATTTGTAATAAATGGTATATCAATAGC 5351
QY 184 ttctcaatcgggaaacagtttttatcgtttgtttattacggagttgacaaaaatctataat 243
Db 5352 TTTAAATATTTTAAAGATATGTACAAAGGGTTTTATTT-----ACAGCCGCCAGAT 5402
QY 244 aaacglaaattcaattatttacttttaaaatggttaataaaaaagattggcggaactgtcgat 303
Db 5403 AAATAGACAAATGTTTTTACAAATACTATGGACTTAAATGAAGTTAATTAATTATTGATA 5462
QY 304 gaacatcagcccgatattattataacatttcccgatgacgctgcgcggaatacaga 363
Db 5463 AAAGAAAGCCAGATTTAATATTATTAAACGTTTCCCTACACAGTTATGTCGGTACTAACT 5522
QY 364 cgcggaactggaagatcattcctacacctcaacggttatgactgattttgtcttcataaa 423
Db 5523 GAGCAATTTTAAACATTAATATTTCCAGTTTGCATGATGATGACAGACTATTCCTTTACATA 5582
QY 424 atttggttcacgaaacggtggataatatattgtgagcagacattacgtggaaggaaaaa 483
Db 5583 AACTGGATTACGCCGTATTTCACAAAGATATTATGTGGCAACAAAGAAACAAACAGAC 5642
QY 484 ctgctggagatcggcactcattccaaagcaatgtaaaaatcacaggaattccaatcagggccg 543
Db 5643 TTCATAGACCTAGGTATTGATCCTTCAACAGTTAAAGTGACAGTATTCCTATTGTATAAC 5702
QY 544 caatttgaagaatcccatgctgctgttgcccgatataaaaaagatacaattttcaccaaac 603
Db 5703 AAATTTGAAACGCCCTATTAAATCAAAAGCAGTGGTTAATAGACAACTTAGATCCAGAT 5762
QY 604 aaaaaagtgcttctgatcattgagcaggtgctcagcgtgtatataaagaacgtaaaagagctg 663
Db 5763 AAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACGATG 5822
QY 664 ---tgcgaaacactgtcgaaggtgacccaagtgcgaagtagttgtcgtgctcgggaaaaat 720
Db 5823 ATTACTGATATATTAGCGAAAGTGCAAATGCACAAAGTAGTTATGATTTTGTGTAAGAGC 5882
QY 721 acggtcttaaaagaatctttgagtcgcttgaagcggaagaaatggtgacaaattaaaagtt 780
Db 5883 AAAGAGCTAAAGCGTCTTTTAAACAGCTAAGTTTAAATTAACGAGAAATGATTGTA---TT 5939
QY 781 ctgggtctatgtgagcgcattgatgagctatttcggaatcacagattgcatgattaccaag 840
Db 5940 CTAGGTATATACCAACACATGAATGAATGATGATGATGATGATGATGATGATGATGATGAT 5999
QY 841 ccgcgagcattactttgacagaagcacacgcatggagtcctgtcattctctgtacaaa 900
Db 6000 CCTGGTGGTATACATAACTGAAGGTTTCGCCGCTGTTGTTTCAATGATTTTCTTAAAT 6059
QY 901 ccgctgctggtggcagaaaaaagaaatgcaaaccttctttgaagaccgagcagtgccatc 960
Db 6060 CCTGCACCTGCTCAAGAGCTTGAATAATGCTTTACTTTTCAAGAAAAAGGTTTTTGGTAAA 6119
QY 961 gttgtgaacgcgtcatgaaga 980
Db 6120 ATCGTGATACTCCAGAAGA 6139
```

RESULT 7

AX016297

LOCUS

AX016297

DEFINITION

Sequence 2 from Patent WO9949052.

AX016297

ACCESSION

AX016297.1

VERSION

GI:10041860

KEYWORDS

Staphylococcus aureus.

ORGANISM

Staphylococcus aureus

Bacteria; Firmicutes; Bacillus/Clostridium group;

Staphylococcus group; Staphylococcus.

REFERENCE

1 (bases 1 to 975)

AUTHORS

Zaeheringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.

TITLE

Processive glycosyltransferase

PAT 07-SEP-2000

QY	883	ctgttcattcgtacaacccgctgcctggcccaggaaagaaaatcgcaacttctttga	942
Db	253906	CRAATTATCATTCACCGAGCCGCTGGTGCATGAGGAACAACCAAAATTTTAATT	253965
QY	943	gaccggagaactgccatcgttgtgaacctcatgaagacatctcagtgcaagtcca	1002
Db	253966	GACGCTGGCCGACGCTAAGGGTAAAGGGGAAGCAAGAATAATTCTTACAACAATCAGCGA	254025
QY	1003	ctcttcgacgatgaagatactcttcgcatcgatgaagaaaaaacattaagaccttcattta	1062
Db	254026	GTTCCTTTACGAGAACGGTGCTTTGGTCCGATCATCAAATCGAGAAAAACTAAA AAAA	254085
QY	1063	gcaaatcctctcgaagtgtatttagagatatcctg	1098
Db	254086	CCAAATGCTGCCAATGAGATTGTGGAGCAGATCGTG	254121

RESULT 9

LOCUS AE001958

DEFINITION Deinococcus radiodurans R1 section 95 of 229 of the complete chromosome 1.

ACCESSION AE001958

VERSION AE0000513

KEYWORDS AE001958.1 GI:6458805

SOURCE Deinococcus radiodurans.

ORGANISM Deinococcus radiodurans

REFERENCE Bacteria: Thermus/Deinococcus group: Deinococcales; Deinococcus.

AUTHORS White,O., Eelsen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M., Vamathevan,J.J., Lam,P., McDonald,I., Utterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et.al.

ECT 22-NOV-1999

TITLE Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1

JOURNAL Science 286 (5444), 1571-1577 (1999)

MEDLINE 20036896

REFERENCE 2 (bases 1 to 13408)

AUTHORS White,O., Eelsen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M., Vamathevan,J.J., Lam,P., McDonald,I., Utterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W., Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S., Smith,H.O., Venter,J.C. and Fraser,C.M.

Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source 1..13408

/organism="Deinococcus radiodurans"

/strain="R1"

/db_xref="taxon:1299"

/chromosome="1"

82..447 /gene="DR1073"

82..447 /gene="DR1073"

/gene="DR1073"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF10658.1"

/db_xref="GI:6458817"

/translation="MTPLLTIVGLSVLLIGMLLSRPKPPOETPIAPRRWLDPA VQAVDFRSSGKTFPVRGQRVWEPRVLILKRDGAEMERLILOQTDLFPASREQLL KIITYDTRCGICTCSGSDA"

431..1186 /gene="DR1074"

431..1186 /gene="DR1074"

/note="similar to SP:P32205 PID:397538 percent identity:


```
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Smith,D.R.
TITLE        1 (bases 1 to 150583)
JOURNAL      Genome Therapeutics Corporation Sequencing Center: Human Genome
SEQUENCE DATA
Unpublished
REFERENCE
AUTHORS      Smith,D.R.
TITLE        2 (bases 1 to 150583)
JOURNAL      Direct Submission
Submitted (25-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE
AUTHORS      Smith,D.R.
TITLE        3 (bases 1 to 150583)
JOURNAL      Direct Submission
Submitted (10-NOV-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT      On Nov 10, 2000 this sequence version replaced gi:9887582.
FEATURES
source
1..150583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RP11-123C3"
/clone_lib="RPCI-11"
BASE COUNT  48699 a 29525 c 28484 g 43875 t
ORIGIN
Query Match      3.9%; Score 44.4; DB 88; Length 150583;
Best Local Similarity 58.2%; Pred. No. 4.2;
Matches 78; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY - 154 gaggtactcaatacctttatttaaaagcttctcaatcggaacagttttatcgtttg 213
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64008 GAGGTATTACTACGCACTACTACAAACATCAACAATTTGGAAATTATATTGAGAA 63949
QY 214 tttattacggaggtgcaaaatcataataaagcgttaattcattcttaaaagt 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63948 ATATGGGATCTACTGCACAAACTATAAATCTTTTAATAAATTTTATGTTAAATTA 63889
QY 274 ggtaataaaagatt 287
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63888 TTAATAAAAAAATT 63875
RESULT 14
AL354684/c
LOCUS      AL354684 170934 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-191021, *** SEQUENCING IN
PROGRESS ***, 17 unordered pieces.
ACCESSION AL354684
VERSION AL354684.4 GI:9863681
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Burton,J.
TITLE        1 (bases 1 to 170934)
JOURNAL      Direct Submission
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT      On Aug 21, 2000 this sequence version replaced gi:9213264.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
```

```
Center project name: ba191021
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 160941 bases at least Q40
Consensus quality: 164931 bases at least Q30
Consensus quality: 167496 bases at least Q20
Insert size: 169334; sum-of-contigs
Insert size: 164439; agarose-fp
Quality coverage: 3.29x in Q20 bases; sum-of-contigs Quality
coverage: 3.70x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11172: contig of 11172 bp in length
* 11173 11272: gap of 100 bp
* 11273 14966: contig of 3694 bp in length
* 14967 15066: gap of 100 bp
* 15067 19307: contig of 4241 bp in length
* 19308 19407: gap of 100 bp
* 19408 63180: contig of 43773 bp in length
* 63181 63280: gap of 100 bp
* 63281 68271: contig of 4991 bp in length
* 68272 68371: gap of 100 bp
* 68372 70384: contig of 2013 bp in length
* 70385 70484: gap of 100 bp
* 70485 73907: contig of 3423 bp in length
* 73908 74007: gap of 100 bp
* 74008 81123: contig of 7116 bp in length
* 81124 81223: gap of 100 bp
* 81224 98475: contig of 17252 bp in length
* 98476 98575: gap of 100 bp
* 98576 101002: contig of 2427 bp in length
* 101003 101102: gap of 100 bp
* 101103 147781: contig of 46679 bp in length
* 147782 147881: gap of 100 bp
* 147882 150311: contig of 2430 bp in length
* 150312 150411: gap of 100 bp
* 150412 153588: contig of 3177 bp in length
* 153589 153688: gap of 100 bp
* 153689 157277: contig of 3589 bp in length
* 157278 157377: gap of 100 bp
* 157378 166456: contig of 9079 bp in length
* 166457 166556: gap of 100 bp
* 166557 168668: contig of 2112 bp in length
* 168669 168768: gap of 100 bp
* 168769 170934: contig of 2166 bp in length.
FEATURES
source
1..170934
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-191021"
/clone_lib="RPCI-11.1"
1..11172
/note="assembly_fragment:00531
fragment_chain:1"
11273..14966
/note="assembly_fragment:01529
fragment_chain:1"
15067..19307
/note="assembly_fragment:00958
fragment_chain:1"
19408..63180
/note="assembly_fragment:00538
fragment_chain:2"
misc_feature
11273..14966
misc_feature
11273..14966
misc_feature
15067..19307
misc_feature
19408..63180
```



```
IFKDSNLAKTNKORRVOIVTAAALQSTHTGNTDQAQKAMGIIITMALITANLHSTANDIP
TWSQFACATAGLGTSGCKWKIIKTGVGKIMKIRPVNGVSADLQGAALIFGATFIHLP
VSTHVISSSLGVGASHRVKGVNMGTAKRISLHSGSSRFRQRHLVSPITLF"
complement(2613. .3230)
/gene="ykaa"
complement(2613. .3230)
/gene="ykaa"
/function="unknown"
/codon_start=1
/transl_table=11
/product="ykaa"
/protein_id="CAA05565.1"
/db_xref="GI:2632006"
/db_xref="SPTREMBL:034454"
/transl_table="MIRKDKFSLTEIAKNIIDEAEYFVNFKVTNQITLKEFADT
LKEVETGDNHVHMILKANKAFITPIEREDILQLTNSLDDVLDGIEHFSAMMEIFS
TSSEHDHDKSCYRECAKEILITIELLAENRLKDIOHPAIKIEYHSDCNLHRKSL
KNLFGNETDPIKVIQYREIYETLEEIADSCQSVANNLETIIMKNA"
complement(3506. .4822)
/gene="ykaa"
complement(3506. .4822)
/gene="ykaa"
/function="unknown"
/codon_start=1
/transl_table=11
/product="ykaa"
/protein_id="CAA05566.1"
/db_xref="GI:2632007"
/db_xref="SPTREMBL:034739"
/transl_table="MHRDNGLKEIGLLEFALTIVIGTIGSGVFMKPGAVLAYSGBS
KMAFLWGLGLTLAGLTIAEIGTQIPKGGTYLLEEVYGEFVGLCGWQIILY
GPAITGALGYFGLMANLFGWSGLSKVIGIIAIVLCVINIIGTIGGFGVQPLTTI
GKLPIACIIVFGLWGDQHIPTAVNESISDMNFGAAILATLFAIDGWIILALGGM
KNPEKLLPRAMTGGLLIVTALYIFINFALLHLSANEIIVTIGENATSTAAITMFGSIG
GKLISVGIIVSIFCLNGKVLSPRVSFMAERKOLPAEKLSHVHPSPFTPTAISF
QIALALIMLISNDKLSLSIEMLYIFYVMAPFAVILKRAKGEKRAYSVPLYPH
PILAIAGSFVGLSTLTDNMSCGLSILIGLAGLPPVYGMKKRKAS"
5211. .6161
/gene="ykaa"
5211. .6161
/gene="ykaa"
/feature="putative dioxygenase"
/codon_start=1
/transl_table=11
/product="ykaa"
/protein_id="CAA05567.1"
/db_xref="GI:2632008"
/db_xref="SPTREMBL:034689"
/transl_table="MKVNGIHHVSALTADAQKNLDFYKVLKLVKKSVMODEPTMY
HLFGDEVANPGTELTTFEIPRIAPFHAGTNSISSIGLRVPGTEALHYKWEREEQVY
THSGISRRAGRDILAFQDHEGQRLVLDEEGKGLPVKVGOSGIEPFRGLGPRV
TVPYAEPPLVHLNLTGFTISREPVEGQTAVILESGEGGAATEIHLIERNDLPRR
QKGSVHVFAPRVDEDELAGWHRIISREGFSNGIVERYFYKALYFREPNGILFELS
TDGPGFVNDLDELQTLIAPPVLEHRRRAIEAKLPQ"
6412. .8562
/gene="ykaa"
6412. .8562
/gene="ykaa"
/function="unknown"
/codon_start=1
/transl_table=11
/product="ykaa"
/protein_id="CAA05568.1"
/db_xref="GI:2632009"
/db_xref="SPTREMBL:034575"
/transl_table="MEKKRELDIVLILLIASAFNIYNIWKDDTVNPPYTAVTSM
MOSPHNFYASFDAAGFTVDKPPVYQIOTISALIFGMHGSVILPOALAGVGSVLL
MYLLIKPTFGKTIARISFVMACPTPIAVARTNNVDALLVPELLATWLLFKAIRG
KLIWLLAAFPVGVGFTKMLQAYMILPAFLFYLIAANATIKKIYVSLVSALAA
VLSUWPLIVNIPASKRPYIGSSQTNVSLFAFYNGIQRITGONSGGGGAPNPKDAS
KEMSSDNTAPPNQSNGSSSDGKSNNGMAAPPNGQSGGGPSPGGGQGG
GPGDGGKGTGSKMQSGSGMGFTGTGTPGLRFQELSDQISWLPFAIFGIAGLL
IAGARERRRLSVEQETVFWVAVLVPVPIAGFFSVAEFPFHYIIMLAPREFAPLVGAGV
```

```
ALVHLRYNOTGKRAWLLPGLAIATTGTFELFRLRNYNDQIGVMSIGVGVGLVLSAIAL
LLPKOROKPSYVYVSLAAILALALVMPYMASTPLLYGNSLSPETGPOLASMSGKMG
MSDATVNEKLIKLENNNSAEYLFPATDSNTAAPPYIIKTKKAVMAIGGYSGSDPALI
LTQFKLVKKGKVKIFLASGMGRGNNDIVIEWEKKEVASEKWKQSSSDOKTENTDS
ADTSSSKASGSKMGKGGPGGNGOSATLYELHADE"
8574. .9545
/gene="ykaa"
8574. .9545
/gene="ykaa"
/feature="putative dolichol phosphate mannose synthetase"
/codon_start=1
/transl_table=11
/product="ykaa"
/protein_id="CAA05569.1"
/db_xref="GI:2632010"
/db_xref="SWISS-PROT:034319"
/transl_table="MSRHIOYVIVPVYNEELVTHETQRLKEYVMDTKENYELLFVN
DGSKDRSIEILREHSLDIPRKIIDFSRNFHQIATAGMDYAGNIAVITADLQDP
PELILEMIEKKEGYEVYAVRTKGETEFKQTAAMFYRLLSGMDTIDIPDTGDF
RLMDRKVCDKMKRLKPNFVGLVSWGSKQTAVEVVRDERLAGETKYPCLKMLKLS
MDGITTESHKPLKILASAGILMSGTGELYMFIVLYLKLFTDSTITGHSSLIIVTOLFS
GIVLLILGIVGEXIGRIYDEAKDRPLIIVOKSYGIENKRLYRDOHMS"
complement(10063. .11412)
/gene="ykaa"
complement(10063. .11412)
/gene="ykaa"
/feature="putative serine protease, heat-shock inducible;
homologous to HtrA from E. coli"
/codon_start=1
/transl_table=11
/product="ykaa"
/protein_id="CAA05570.1"
/db_xref="GI:2632011"
/db_xref="SPTREMBL:034358"
/transl_table="MDNVRDNRRTKGNNEVELTKENDQASYSARNVVIHDOEKKRG
FGWFRPLGVIGSLGIYITFTPLGNHDSQDTAKQSSSQDTQSTYATSTSESKK
SSSSSAFKSDEDSKISDMVEDLSPAIIVITNQAQNSSLFGSSSSSDESDTESGSG"
Query Match 3.8%; Score 44.2; DB 2; Length 55593;
Best Local Similarity 54.7%; Pred. No. 4.3;
Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
Qy 392 tcaacgttatgactgattttgtcttataaaatttgggttcacgaaacgtgataaat 451
Db 51832 TGAATGTATACACTGATTTTTCGTGAATCAATTATGGGACGGAACATTCATTACC 51773
Qy 452 attatgtggcgacagattacgtgaaggaaacgtgctgagatcggcactcatcaagca 511
Db 51772 ATTTGTTCGACGACGCAAGTAAGAAGCAGCTAATATCAGAGGTTATCGATCAGAACAA 51713
Qy 512 atgtaaaaatcacagggaattccaatcagggccgcaatttgaa 552
Db 51712 ACATATACTTGACCGGCATTCCAGTTCACCAAAATTTTGAA 51672
```

Search completed: June 29, 2001, 11:53:36
Job time: 8429 sec

7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:57:18 ; Search time 28.15 Seconds
(without alignments)
1033.701 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILTANYGNHGVQ.....EDILKESEMMTAKQAKVLS 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1970	100.0	382	2 C69935	cell wall synthesis
2	368	18.7	374	2 E83894	hypothetical prote
3	353	17.9	411	2 F75439	probable cell wall
4	317	16.1	464	2 E84499	probable monolac
5	309	15.7	373	2 B69860	conserved hypothet
6	303.5	15.4	525	2 T10478	probable 1,2-diacy
7	301.5	15.3	533	2 T05092	probable 1,2-diacy
8	293	14.9	468	2 T52269	1,2-diacylglycerol
9	196	9.9	363	2 JC1275	phospho-N-acetylmu
10	195	9.9	344	2 C70401	phospho-N-acetylmu
11	178	9.0	363	2 E83970	UDP-N-acetylglucos
12	171	8.7	351	2 D64185	UDP-N-acetylglucos
13	170.5	8.7	363	2 F70195	UDP-N-acetylglucos
14	163	8.3	383	2 E70156	lipopolysaccharide
15	159	8.1	357	2 E86823	peptidoglycan synt
16	156.5	7.9	398	2 F64456	hypothetical prote
17	138	7.0	378	2 A72304	UDP-N-acetylglucos
18	137.5	7.0	364	2 T34954	probable UDP-N-ace
19	136.5	6.9	458	2 C71420	hypothetical prote
20	136	6.9	353	2 C64664	transferase, pepti
21	135	6.9	484	2 E64432	spore coat polysac
22	134	6.8	353	2 G71852	udp-n-acetylglucos
23	130.5	6.6	392	1 C69851	macrolide glycosyl
24	128.5	6.5	410	2 E70579	probable murG prot
25	128	6.5	406	2 C72340	probable hexosyltr
26	123.5	6.3	367	2 D82763	UDP-N-acetylglucos
27	123	6.2	355	1 BVSCMG	UDP-N-acetylglucos
28	123	6.2	355	2 E85491	hypothetical prote
29	122.5	6.2	339	2 B72402	UDP-N-acetylglucos

30	122.5	6.2	342	2 D81306	probable UDP-N-ace
31	121.5	6.2	385	2 C71699	murG protein (murG
32	119	6.0	418	2 G75496	UDP-N-acetylglucos
33	119	6.0	478	2 T08395	UDP-glucose glucos
34	117.5	6.0	380	2 D70049	UDP-N-acetylglucos
35	117	5.9	400	1 F69142	probable hexosyltr
36	116.5	5.9	338	2 E71699	capM protein (capM
37	116.5	5.9	366	1 D70351	probable hexosyltr
38	116	5.9	363	1 C72590	probable hexosyltr
39	116	5.9	379	2 E86651	LPS biosynthesis p
40	115.5	5.9	571	2 B64469	hypothetical prote
41	114.5	5.8	362	2 B86783	UDP-N-acetylglucos
42	114.5	5.8	404	1 S77553	probable hexosyltr
43	113.5	5.8	354	2 D84955	hypothetical prote
44	113	5.7	371	2 S78863	hypothetical prote
45	112.5	5.7	1068	1 JQ1329	sucrose-phosphate

ALIGNMENTS

RESULT 1
C69935
cell wall synthesis homolog yfpf - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: C69935
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Buschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerman, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toqnoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: C69935
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-382 <KUN>
A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14110.1; PID:el1836
A:Experimental source: strain 168
C:Genetics:
A:Gene: yfpf

Query Match		100.0%	Score 1970;	DB 2;	Length 382;
Best Local Similarity		100.0%	Pred. No. 2.4e-127;		
Matches 382;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MNTNKRVLILTANYGNHGVQVAKTLYEQCVRLGFQHVTVSNLYQESNPVSEVTQYLYLK	60		
Db	1	MNTNKRVLILTANYGNHGVQVAKTLYEQCVRLGFQHVTVSNLYQESNPVSEVTQYLYLK	60		
QY	61	SFSGIKQFYRLFYGVVDKTYNKRKFNIFYFKMGNKRGELVDEHQPDIIINTFFMIVVPEY	120		
Db	61	SFSGIKQFYRLFYGVVDKTYNKRKFNIFYFKMGNKRGELVDEHQPDIIINTFFMIVVPEY	120		
QY	121	RRRTGRVITPFNVTDFCLHKIWHNVNDKYVATDVVKEKLEIGHPSNVKITGPIR	180		
Db	121	RRRTGRVITPFNVTDFCLHKIWHNVNDKYVATDVVKEKLEIGHPSNVKITGPIR	180		
QY	181	POFEESMPVGPYIKKYKYNLSPNKKVLLIMAGHGVKLVNKCENLVKDDQVQVWVCGKN	240		
Db	181	POFEESMPVGPYIKKYKYNLSPNKKVLLIMAGHGVKLVNKCENLVKDDQVQVWVCGKN	240		

, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75439
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <WHI>
A:Cross-references: GB:AE001958; GB:AE001958; GB:AE000513; NID:g6458805; PIDN:AAF10649.1; PID:g6458805
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1076
A:Map position: 1

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83894

	Query Match	17.9%;	Score 353;	DB 2;	Length 411;
	Best Local Similarity	26.2%;	Pred. No. 8.7e-17;		
	Matches 101;	Conservative 77;	Mismatches 164;	Indels 44;	Gaps 11;
QY	6	RVLILTANVGNHGVAKTLYEQCVRLFGQHVTVSNLYQESNPIVSEVTOYL--YLKSPS	63		
Db	29	RALFMSVSLGAGHDQA-----QQAVQKFAERGVGLGAEDHSV-----EYLSTERSFT	78		
QY	64	I-----GKQFYRLFYGVGDKIYNKRKFNIYKEM-----GNKRLGELVDEHPDIIIN	110		
Db	79	VDIYEFELRYAPWLYRGFWLTQQ-----DQPNWISIRFTWLGMAFKDELRURPEVIN	135		
QY	111	TF--PMIVPEYRRRTGRVPTENVMTDFCLHKIWIWHENVDKYVYATDYVYKELLEIGTH	168		
Db	136	SWAPAAVCDTTLTAQGTGREFNCLITVDYBAHIHWARFETDILMVASFEFTRMVLRGV	195		

QY
109 PNVKILGIPKRPQESMPV-----GPIKYKNLSFNKKRVLLIMAGAHGVLLKNVKEL 222
Db 196 PEQVEVTGIPSPAERFVLAADRWAALRAELFSEMSLRGPVPLLLLSGGRGCHYAAADV 255

```

223 ENLVR-DDUQVVWVCCKNFKALSESALSAENGDKLVIGYVERIDELPRIDCMITPK 281
      |||:|      ||:      ||:      |      :||:|
256 TELGNLGRAVQVLVPASROGEGTETI-----GGATVHHLGFRDLPRLLAASDLVVGKA 309
      |||:|      ||:      ||:      ||:      ||:      ||:      ||:      ||:
282 GGTTLTEATAIGVPVILYKPVPGOERENANFFEDRGAIVVNRHEETILESVTSLADEDT 341
      |||:|      |||:|      |||:|      |||:|      |||:|      ||:      ||:
310 GGLTVAEATLGVPLIYYIPGQEEHNADFLEHGGAGLWARARHVRPLVLRAL-DPAE 368
      ||:      ||:      ||:      ||:      ||:      ||:      ||:      ||:
342 LHRMKNKIKDLHLNANSEVILEDILK 367
      ||:      ||:      ||:      ||:      ||:
369 HARLSAGARAVGIPDAADRVRAGAILR 394
      ||:      ||:      ||:      ||:      ||:

RESULT      4
C84499
probable monogalactosyldiacylglycerol synthase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84499
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB84420; MUID:20983487
A:Accession: C84499
A>Status: preliminary
A:Molecule type: DNA

```

Cross-references: GB:AE002933; NID:94734000; PIDN:RAD28676.1; GSPDB:GN00139
C-Genetics:
A:gene: At2g11810
A:Map position: 2
Query Match 16.1%; Score 317; DB 2; Length 464;
Best Local Similarity 25.6%; Pred. No 3e-14;

Query Match	16.1%;	Score 317;	DB 2;	Length 464;
Best Local Similarity	25.6%;	Pred. No. 3e-14;		

Matches	101;	Conservative	90;	Mismatches	171;	Indels	32;	Gaps	16;
---------	------	--------------	-----	------------	------	--------	-----	------	-----

```

5  KRVLIIITANYNGHVOVAKTLEQCQVRGLG---QHVTYVSNLYQESN--PIVSEVTOYIYL 59
   | | | | | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
72  KTVLIIMSTGGGHRASAEAI-RDAFCIBFGDDYRIIHKDVKWKEYTGWPLNDMERQYKFM 130
   | | | | | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
60  KSFESICKQVRFLFYGVGD-KIYNKRKFNIYFKMGKNKRGELVDEHQPDIIINTFFM--- 114
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
131  VK-HVG--LWSVAFHGTSRKHYSLSALAAYYAKEBAGLMEYKPDIIISVHPLMOHI 18
   | | | | | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
115 -IYVPEYRRRTGRVITFFNMDF-CLHKIYWHENVDKYIVATDVVKKLLIEIGTHPSNV 172
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
188  PLAWMKWGLHKKVI-FVTVITDLNTCHRTWFHGHGVSRCYCSKEVAKRALVDGLDDSQI 246
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
173  KITGIPIRQFERS-MPVGPIYKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLYKD 228
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
247  RVFGLPVRSPFPTIILNKNLKELEIDILNLPAILMGGEGGVPQKTALALGDSLNS 306
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
229  DOV---QVVVCGKTKALKESISALEAENGDKLVGYVERIDELFTDCMTKPGGI 284
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
307  KESNPITGLIVICGRNKLVIASLTLSHEWK--TPVKVRFETQMEKMGACDCIITKAGP 364
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
285  TLTEATAIGVPVILYKVPVQKEKANANFPEDRGAAIVNVRHREIILESVTSLIA-DEDTIH 343
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
365  TIAEALICGLPIILNDYIPQKEGNVPYVVDNGAGVFTSPRETAKIYADVFSNKEELK 424
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
344  RMKKNKTDLHLANSSEVILEDILKSEMMTAKOK 377
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
425  KMSENA--LKLKSQEAUV--DIVDKDTHHLSQOQQ 454
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 5
B9860
conserved hypothetical protein yk0N - Bacillus subtilis
C:Species: Bacillus subtilis
C:date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

Query Match 15.7%; Score 309; DB 2; Length 373;
Best Local Similarity 24.6%; Pred. No. 7.7e-14;
Matches 99; Conservative 89; Mismatches 158; Indels 56; Gaps 16;

QY	5	KRVLIIT-ANYGNHGHVQAKTLYEQCVRLGFOHVTVSNLYQESNPITSEVTOYLYLKSPS	63
Db	2	KNILIFPFLSISGTHHHVADALQAELESQGLAAEKI-DIFSYSYRLEKLSVAVYIKWI-	59
QY	64	IGQVRLFLPYGVVDKI-----YNKRKFNIYFKMGKRLGELVDEHQDPIIINT--FPM	114

[illegible]

RESULT 6
Tl0478
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) precursor, chl
N;Alternate names: monogalactosyldiacylglycerol synthase
C;Species: Cucumis sativus (cucumber)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: Tl0478
F;Shimojima, M.; Ohta, H.; Iwamatsu, A.; Masuda, T.; Shioi, Y.; Takamiya, K.
Proc. Natl. Acad. Sci. U.S.A. 94, 333-337, 1997
A;Title: Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolution
A;Reference number: Z17042; MUID:9714442
A;Accession: Tl0478
A;Status: 'preliminary'; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-525 <SHI>
A;Cross-references: EMBL:U62622; NID:g1805253; PIDN:AAC49624.1; PID:g1805254
A;Experimental source: cv. Anonagajimai; 5 day old seedlings
C;Genetics:

Query Match	15.4%	Score 303.5	DB 2	Length 525
Best Local Similarity	25.7%	Pred. NO. 2.9e-13		
Matches 100	Conservative	77	Mismatches 179	Indels 33
				Gaps 13

```

Qy 5 KRVLIIPANTYNGNH---VOVAKTLYEQCVRLGFQHFVTSNLYQESNPVIVSEVTQYLYLKS 61
Db 136 KRVLIISDITGGHRSAAEAIKAAFNEEFQNNYQ-VFITDLWTDHTPWPENQLPRSYNFL 194
Qy 62 FSIQKQFVRLFYGVGDYKIYKNKRFNIVFKMGKRLGLVDHOPDIIINFPM-----IV 116
Db 195 VKHGTLKMYIYVYAPKVIHQSNFAATSTFIAREVAKGLMKYRPDIISVHPLMQHVPIR 254
Qy 117 VPEYRRRTGRVIFTFNWMTDF-CLHKIWHENVDKYVATDYVKKEKLELTGTHPSNVKIT 175
Db 255 ILRSKGILLNKIVFT-TVVTDLSTCHPTWFHKLTRCYCPSTEVAKRALTAGLOPSKLKVF 313
Qy 176 GIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGHG---VLKNVKLELCENLVKDD 229
Db 314 GLPVRPSFVK--PIRPKIELRKLGHMDENLPAVLLMGCGEGMGPIEATAKALSALYDEN 371
Qy 230 Q-----VOVVVVCNKNTALKESLSALEAENGDKLVGLYVERIDELFRITDCMTTKPGGIT 285
Db 372 HGEPFGVLVLCIGHNKKLAGRLRSIDWK--VPVQVGFVTKMEECMGACDCITTKRAGPG 429

```

QY 286 LREATAIGVPVILYKVPVQGEKENANFFEDRGAAIIVNRRHEILLESVTSILLADE-DTLHR 344
Db 430 IAEAMIRGLPIILINDYIAGQACGNVPYVVGCGCKSKSPKEIANIVAKWFGPKADELLI 489

QY 345 MKRNIKDLHLANSSEVI-----LEDILKE 368

Db 490 MSQNA--LRLARPDVAFKIVHDLHLVKQ 516

RESULT 7

T05092
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) - Arabidopsis th

N;Alternate names: Protein F28M20.30
C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999

C;Accession: T05092

R;Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hohelsel, J.; Mewes,
submitted to the Protein Sequence Database, November 1998

A;Reference number: Z15398

A;Accession: T05092

A;Molecule type: BEV

A;Residues: 1-533 <BEV>

A;Cross-references: EMBL:AL031004

A;Experimental source: cultivar Columbia; BAC clone F28M20

C;Genetics:

A;Map position: 4

A;Introns: 175/3; 233/2; 287/2; 334/3; 409/3; 430/3; 455/3

A;Note: F28M20.30

C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 15.3%; Score 301.5; DB'2; Length 533;
Best Local Similarity 24.8%; Pred. No. 4.1e-13;
Matches 101; Conservative 86; Mismatches 172; Indels 49; Gaps 17;

QY 5 KRVLILTANYNGHVQVAKTYLQCVRLGFQ-----HVTNVLQESN--PIVSEVTQ 55

Db 142 KKVILMSDTGGGHRASA-----EAIAPNQEFGEYQVFIITDLTHTPWPFPNQLPRS 196

QY 56 YLYLKSISGKQFYRLFYGV--KIYKRNKFNLYFKMGNKRLGELVDEHOPDIIITFPM 114

Db 197 YNFIKKG---TLWKMTYIGTSPIRVHQSNFATSTFIARIAGLKKYQPDIIISVHPL 253

QY 115 I--VVPYRRRTG--RVIPTFNVTDF--CLHKIWHVHNDKYVATDYVYKEKLLLEIGTHP 169

Db 254 MGVPLRLVRSKGLLKKIVFTTITDLSCHPTWFKHLVTRCYCPSPEVAKRAQKAGLET 313

QY 170 SNVITGIPRPQFEESMPVGP---YKKNLSPNKKVLLIMAGAHV--LKNVKELCEN 224

Db 314 SQIKVYGLVPRPSFK--VPRPKVELRELGMENLPAVLLMGCGMGPIEATARALAD 371

QY 225 LVKDDQV-----QVWVCGKNTALKESLSALEANGDKLVGVVERIDELFRITDCMIT 279

Db 372 ALYDNKGEAVQVLIICGRNKKLSLSDWK--IPVQVKGFTKMECGACDCIIT 429

QY 280 KPGGITLTAIPAIGVPVILYKVPVQGEKENANFFEDRGAAIIVNRRHEILLESVTSILLAD- 338

Db 430 KAGPTIAEAMIRGLPIILNGYIAGQACGNVPYVVGCG-----GKFSKSPKEISKIVADW 485

QY 339 ----EDTLHRMKNKDLHLANSSEVILEDLKESEMMTAKQAKVLS 382

Db 486 FGPASKELEIMSONA--LRLA-KPEAVFKIVHDMHVELVRKNLSLPQLS 530

RESULT 8

T52269
1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) [imported] - Arabidopsis th

N;Alternate names: monogalactosyldiacylglycerol synthase

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000

C;Accession: T52269

R;Awai, K.; Shimojima, M.; Masuda, T.; Takamiya, K.I.; Ohta, H.

submitted to the EMBL Data Library, July 1998

A;Description: cDNA cloning of an Arabidopsis monogalactosyldiacylglycerol synthase r

A;Reference number: Z26011

A;Accession: T52269

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-468 <AWA>

A;Cross-references: EMBL:AJ000331; PIDN:CAA04005.1

A;Experimental source: cultivar Columbia

C;Genetics:

A;Gene: mgd

C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 14.9%; Score 293; DB 2; Length 468;

Best Local Similarity 25.1%; Pred. No. 1.3e-12;

Matches 98; Conservative 86; Mismatches 172; Indels 34; Gaps 14;

QY 5 KRVLILTANYNGHVQVAKTYLQEC--VRLGFOH-VTVSNLYQESN--PIVSEVTQYLYLK 60

Db 69 KKVILMSDTGGGHRASAEAIRDAFKIEGDKYRVIVKDVWKEYTGWPLNDMERSYKFMV 128

QY 61 SPSIGKQFYRLFYGV-----VDKIYKRNKFNLYFKMGNKRLGELVDEHOPDIIITFPM- 114

Db 129 KH---VOLMKVAFSTSPKWIHSCLAAATAAYAKEVEAGL-----MEYKPEIIISVHPLM 181

QY 115 ----IWPYRRRTGRTVPTFNVTDF--CLHKIWHVHNDKYVATDYVYKEKLLLEIGTHP 169

Db 182 QHPLVWLAKQELQKRVL--FVTVITDLNTCHPTWFKHPGNYRCYQEQVAKRALFDGLDE 240

QY 170 SNVITGIPRPQFEESMPV--GPIYKKNLSPNKKVLLIMAGAHV--LKNVKELCENLV 226

Db 241 SOVRVGLVPRPSFARAVLVKDDLRKELEMDQDLRAVLLMGCGMGVPYKETALEEF 300

QY 227 KDDQ-----QVWVCGKNTALKESLSALEANGDKLVGVVERIDELFRITDCMITRP 281

Db 301 YKXENKPTGQVWVIGCRNKKLASALEADWK--IPVKGFTQMEKMGACDCIITRA 358

QY 282 GGTTLTAIPAIGVPVILYKVPVQGEKENANFFEDRGAAIIVNRRHEILLESVTSLLADE-D 340

Db 359 GPGTIAESLIRSLPIILNDYIPQGEKGNVPYVVGAGVFTRSPKETARIVGEWFSKTD 418

QY 341 TLHRMKNKDLHLANSSEVILEDLKESE 370

Db 419 ELEQTSNARKLAQPEAVFDIVADIDELSE 448

RESULT 9

JC1275

phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) - Bacillus subtilis
N;Alternate names: murG protein; UPD-N-acetylglucosamine--N-acetylmuramyl-(pentapepti

C;Species: Bacillus subtilis

C;Date: 30-Sep-1993 #sequence_revision 06-Feb-1995 #text_change 20-Jun-2000

C;Accession: S25763; JC1275; S26499; G69662; S22212

R;Henriques, A.; de Lencastre, H.; Piggot, P.

Biochimie 74, 735-748, 1992

A;Title: A Bacillus subtilis morphogene cluster that includes spoVE is homologous to

A;Reference number: S25762; MUID:93003529

A;Accession: S25763

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-363 <HEN>

A;Cross-references: EMBL:X64259; NID:g39994; PIDN:CAA45558.1; PID:g39995

R;Miyao, A.; Yoshimura, A.; Sato, T.; Yamamoto, T.; Theeragool, G.; Kobayashi, Y.

Gene 118, 147-148, 1992

A;Title: Sequence of the Bacillus subtilis homolog of the Escherichia coli cell-divis

A;Reference number: JC1275; MUID:92380484

A;Accession: JC1275

A;Molecule type: DNA

A;Residues: 1-180, 'G', 182-269, 'T', 271-363 <MIY>

A;Cross-references: GB:DI0602; DBJ:J090538; NID:g216299; PIDN:BAA01454.1; PID:g216300

R;Beall, B.; Lutkenhaus, J.

J. Bacteriol. 171, 6821-6834, 1989

Query Match 9.0%; Score 178; DB 2; Length 363;
Best Local Similarity 23.1%; Pred. No. 6.7e-05;
Matches 72; Conservative 62; Mismatches 124; Indels 54; Gaps
Qy 87 IYFKMGNKRGELVDEHQDPIIINTFPMIVPEYRRRTGRVITPTNNWTFDFCLHKIWVHE 146

RESULT 10
C70401
phospho-N'-acetylmutamoyl-pentapeptide transferase - Aquifex aeolicus
C/Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1998

Db 75 VRFLGTRKAKALNEKPDVIGTGYVCPVYAAAKLKIPV-----IHE 122
Qy 147 N-----VDKY---YVATDYVKEKLEIGHPSNVKITGIPRQF-----EESMPVGPI 192
Db 123 QNSVPGLTNKLRSYVDRIAICFEAEAFPPKNVFTGNPRASEVMSGNREGL-----177
Qy 193 YKKYNLSPNKKVLLIMAGHGVKNVKELCNLYKDDQV---QVVVCG--KNTALKESL 247
Db 178 -RSLGIRKPNKTVLIVGSGRGA-RPINDAFMSILSDVAKAPQFVYVTVGVHVERVOEQM 235
Qy 248 SAL-EAENGDKLKVGVVERIDELFRITDCMITKPGGTTITATEAIGVPPVILYKPPV---303
Db 236 KSIQOPEN---VIVQPFTHNMPDVSADVLIVARAGATTLAEITFALGLPSILI-PSPYVT 291
Qy 304 -GQKENANFFEDGAAIVNRHE-----EILESVTSLADEDTLHRMKNKIKDLHLANS 358
Db 292 NNHOEKNAALSKDAAILRKESELTGRLEDDIDIMVTPGRDLAMQAAKALGVPTAA 351
Qy 359 E---VILEDILK 367
Db 352 EKLHMLVKEVAK 363
RESULT 12
D64185
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-ac
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 26-Aug-1999
C:Accession: D64185
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 456-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: D64185
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-351 <TIGR>
A:Cross-references: GB:U32793; GB:U42023; NID:g1574683; PIDN:AAC22793.1; PID:g1574693; T
C:Genetics:
C:Gene: murG
C:Function:
A:Pathway: peptidoglycan biosynthesis
C:Superfamily: murG protein
C:Keywords: cell division; cell wall; glycosyltransferase; hexosyltransferase; peptidogl

Query Match 8.7%; Score 171; DB 2; Length 351;
Best Local Similarity 23.9%; Pred. No. 0.00019;
Matches 95; Conservative 72; Mismatches 150; Indels 80; Gaps 22;
Qy 4 NKRVLILFANYGNHV-----QVAKTYEQ---CVRLGFQHVTVSNLYQESNPVSEVTVQY 56
Db 3 NKKLLVMAGGTG-GHVFPAIAVAOTLQKQENDICWLG----TKDRMEAGLVPKYGIPIRF 57
Qy 57 LVYKFSIGKQFVRLFYGVGVYKRNKFNFIYKMGKRLGELVDEHQPDIINTFFMIV 116
Db 58 IQISGLR-QK-----GIKALN-APAIAPRAVLQAK--KIQEKPDVAVLG-----99
Qy 117 VPEYRRRTGRVITPNNVMTDFCLHKIWHNVENYKDYVATDYVKEKLL-EIGT-----HPS 170
Db 100 -----MGYVSGPAGVAAKLCGPVILHEQ-----NAIAGLTNKLGLKIACTVQLQAFPT 148
Qy 171 ---NVKITGIPRQPEESMPVGPYKKNLSPNKKVLLIMAGA-----HGVLKNVKE 220
Db 149 APFAEVGNPNVREDLFE-MPNPDI--RFSDEEKLRLVLVGGSGARVNLNITLKPVAQ 205
Qy 221 LGENLVKDDQVQVVVCGKNTALKESLSALEAENGDKLVGLYVERIDELFRITDCMITK 280
Db 206 LADKLEFRHQV-----GKGAIV-EVQGLYENLEQVKITFEIDNMAEAYAWADWICR 257

Qy 281 PGGITLATEAIGVPPVILYKPPVGOEKE-----NANFEEDGAAIVVNRHE-----EILESV 332
Db 258 SGALTVCETAAVGAATAIF---VPQHKDRQOYLNAKILSDVGAAKIIEQADLTPEILVNY 314
Qy 333 TSLADEDTLHRMKNKIKDLHLANSSEVILEDTIKES 369
Db 315 LKNLTRENLL-QMALKAKTMSMPNAQRAEVIKQYS 350
RESULT 13
F70195
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 24-Nov-1999
C:Accession: F70195
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: F70195
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-363 <KLE>
A:Cross-references: GB:AE001176; GB:AE000783; NID:g2688699; PIDN:AAC67113.1; PID:g268
A:Experimental source: strain B31
C:Superfamily: murG protein

Query Match 8.7%; Score 170.5; DB 2; Length 363;
Best Local Similarity 20.6%; Pred. No. 0.00022;
Matches 82; Conservative 84; Mismatches 171; Indels 61; Gaps 16;
Qy 3 TNKRVLLITANYGNHV-----QVAKTYEQCVRLGFQHVTVSNLYQESNPVSEVTVQYLY 58
Db 2 SNKKIIFFTGGTGGVHPGIGISIIQKLKEFDNEIEFFWICKKNSIEEK--LIKEQDNKIF 59
Qy 59 LKPSIGKQFVRLFYGVGVYKKNKFNFIYK--MGNKRLGELVDEHQPDIINTFFMIV 116
Db 60 I-SIPCGK-----LRRVFSFKNETDFKVLGIKISFYLLKPKYKQIIVATGCVFS 109
Qy 117 VPEYRRRTGRVITPNNVMTDFCLHKIWHNE-----NVDKYVATDYVKEKLE-IG 166
Db 110 TPAI-----IASSLLKIKSITHMDLPGLATKINSFANNIHISFESEKIF 157
Qy 167 THPSNVKITGIPRQPEESMPVGPYKKNLSPNKKVLLIMAGAAG--VLKNVKELCEN 224
Db 158 KNYKNIIYTGSPIRREFLNPD--KIQLQNTQNTKPIIISILGSGIGANALNL-ALC-- 212
Qy 225 LVKDDQVQVVVCGKNTALKESLSALEAENGDKLVGLYVERIDELFRITDCMITKPGGI 284
Db 213 IKKDAEIVFIHQSKN-----LNDLSEKNYLRROFFN-AEEMASIVKFSNLIISRAGAG 265
Qy 285 TLTAATAIGVDPVILY---KPVGQKEKNANFEEDGAAIVVNRHE-----EILESVTSLA 337
Db 266 AIKEFANAGACAIIIPPKGSRGDOIKNAKLLTNONACIYIDEDILNINILKIITKTLK 325
Qy 338 DEDTLHRMKNKIKDLHLANSSEVILEDTIKESMMTAK 375
Db 326 DREKINSLEKNKFNKNKSHSTLAKLIIKIDIKETSK 363

RESULT 14
E70156
lipopolysaccharide biosynthesis-related protein homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 17-Mar-2000
C:Accession: E70156
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
A:Reference number: A70100; MUID:98065943
A:Accession: E70156
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-383 <KLE>
A:Cross-references: GB:AE001150; GB:AE000783; NID:g2688359; PIDN:AA06815.1; PID:g268836
A:Experimental source: strain B31
C:Superfamily: probable hexosyltransferase ytxN

Query Match 8.3%; Score 163; DB 2; Length 383;
Best Local Similarity 22.7%; Pred. No. 0.00076;
Matches 80; Conservative 70; Mismatches 130; Indels 72; Gaps 21;
QY 74 YGVDKIYKRNKFNIFKMGKRR-LGELVDEHPDII--INTFPMIVVPEYRRRTGR--- 126
DB 52 YRCSSIQINKKLDAVIAFPNKRKISKIIQSYKPDIIHTHSEFSM-----GKIGKQIAL 104
QY 127 -----VPTFNMTDFCLHIWV-----HENVDKYIV-ATDVVKEKLEIGT 167
DB 105 KHNPIVHTSHTMWDYYLHGLIFKYPFKPKDKMRKHYNKIKHFYPPSSKAKERYFQLSN 164
QY 168 HPSNVKI--TGIPRPOFEESM--PVGPIYKYNLSPNKKVLLIMAGAHGVKLVKELC 222
DB 165 NSSNYKIIPNGVD-RKLFITLSKEKDEILKHNKIQOTDKI-IIFVGRINKERNINLV 222
QY 223 ENLVKDDQVO---VVVCGKNTALKESLSALEAENG--DKLVKLVGVV--ERIDELFRIT 274
DB 223 THL-KLLMQNNYKLLIGCKSGSEEK-IGNFSIKHGLEKQILLIGTIPWEEIYVYKIS 280
QY 275 DCMITKPGG-----ITLTAETAIGVPVIL-----YKVPVGOEKENANFFEDRGAIVVNRH 325
DB 281 DIFASLSKSEVPMTVIEALTAGIPAILINDYTKDV---IKEGINGF-----LIKKY 330
QY 326 EEILESVTSLADEDTLHRMKNKIKDLHLANSSEVILEDLKE--SEMPTAK 375
DB 331 ENLSRYIDKVIKDEILKKEKENAKK-HSTKFSYFFTKKIKNYSEIARK 381

RESULT 15
E86823
peptidoglycan synthesis protein MurG [imported] - Lactococcus lactis subsp. lactis (stra
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86823
R:Boletini, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrh
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: E86823
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <STO>
A:Cross-references: GB:AE005176; NID:gl2724594; PIDN:AA05687.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 8.1%; Score 159; DB 2; Length 357;
Best Local Similarity 20.9%; Pred. No. 0.0013;
Matches 72; Conservative 51; Mismatches 129; Indels 92; Gaps 13;
QY 54 TOYLKLSFSIGKQFVRLFYGVYDKYIKRKNFYKMGKRLGELVDEHPDIIINTFP 113
DB 73 TAYKFKSVSDAK-----KIMKFKPDVVLGTGG 101
QY 114 MIVVPEYRRRTGRVITFNWTFCLHKIHWENVNDKYIVATDVVKEKL--LEIGH--- 168

DB 102 YVAGPVVYAAQLKIPTI-----IHEGNSFPGITNRLAKKVDRIAGVGHAAE 149
QY 169 -----PSNVKITGIPRQFEESMPVGIYKYNLSPNKKVLLIMAGAHGVKLVKELCEN 224
DB 150 QYFPASKTTTFTG---NPRAQEVADAAQAQVEKF-----EPTVWIFGSGRALKLNNAFIEA 202
QY 225 L--VKDDQVOVVVVCGR--NTALKESLSALEAENGDKLVGVYVERIDELFRITDCMITK 280
DB 203 LPELAQRSFTVIVASGEIYYDDYKETFN--OYKENSNLDIRPYINNNTELLAKSQLEFLGR 260
QY 281 PGGITITEATAIGVPVILYKPVPG-----QEKENANFFEDRGAIVVNRHHEEILLESVTSLL 336
DB 261 SGSTTIAEVTALGLPAV-YVPSPNVTADQOTKNAQEVVDGAAIIK----- 306
QY 337 ADEDTLHRMKNKIKDLHLANSSEVILEDLKESEMPTAKOKAKV 380
DB 307 -DED-----LTGOTLVEAISNILENNEKYQEMQAASLKAGV 341

Search completed: June 29, 2001, 09:00:19
Job time: 181 sec

7

QY 181 POFESMPVPIYKYNLSPNKKVLLIMAGAHGVLKNVLCENLVKDDQVQVVVCGKN 240
DB 181 POFESMPVPIYKYNLSPNKKVLLIMAGAHGVLKNVLCENLVKDDQVQVVVCGKN 240
QY 241 TALRESLSALEAENGDKLVGYVERIDELFRITDCMITKPGGTTLTAATAIGVVPVILYK 300
DB 241 TALRESLSALEAENGDKLVGYVERIDELFRITDCMITKPGGTTLTAATAIGVVPVILYK 300
QY 301 PVPQGEKENAFFEDRGAAIYVNNHEEILSVTSLLADEDTLHRMKNKIKDLHLANSSEV 360
DB 301 PVPQGEKENAFFEDRGAAIYVNNHEEILSVTSLLADEDTLHRMKNKIKDLHLANSSEV 360
QY 361 ILEDILKESEMNTAKOKAKVLS 382
DB 361 ILEDILKESEMNTAKOKAKVLS 382

RESULT 2

MURG_BACSU STANDARD; PRT; 363 AA.
AC P37585; P18578; 059247;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGUCOSAMINE TRANSFERASE
DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGCLNAC GLCNAC TRANSFERASE).
GN MURG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92380484; PubMed=1387377;
RA Miyao A., Yoshimura A., Sato T., Yamamoto T., Theeragool G., Kobayashi Y.;
RT "Sequence of the Bacillus subtilis homolog of the Escherichia coli cell-division gene murg."
RL Gene 118:147-148(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Henriques A.O., de Lencastre H., Pigot P.J.;
RT "A Bacillus subtilis morphogene cluster that includes spore is homologous to the mra region of Escherichia coli."
RL Biochimie 74:735-748(1992).
RN [3]
RP SEQUENCE OF 216-363 FROM N.A.
RX MEDLINE=90078133; PubMed=2556375;
RA Beall B., Lutkenhaus J.;
RT "Nucleotide sequence and insertional inactivation of a Bacillus subtilis gene that affects cell division, sporulation, and temperature sensitivity."
RL J. Bacteriol. 171:6821-6834(1989).
CC -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-(PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II).
CC -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC

DR EMBL; D10602; BAA01454.1; -
DR EMBL; X64259; CAA45558.1; -
DR EMBL; M31827; AAB83968.1; -
DR EMBL; Z99111; CAB13395.1; -
DR PIR; JCI275; JCI275.
DR PIR; S26499; S26499.
DR Subtilist; BG1027; murg.
KW Transferase; Glycosyltransferase; Cell division; Cell wall;
KW Membrane; Peptidoglycan synthesis.
FT CONFLICT 181 K -> G (IN REF. 1).
FT CONFLICT 270 A -> T (IN REF. 1).
SQ SEQUENCE 363 AA; 39977 MW; 185B20688C732489 CRC64;

Query Match 9.9%; Score 196; DB 1; Length 363;
Best Local Similarity 23.1%; Pred. No. 1.4e-06;
Matches 93; Conservative 76; Mismatches 153; Indels 80; Gaps 20;

QY 6 RVILITANYGNHGVAKTLYEQCVRLGFQHVTVSNLY-----QESNPIVS- 51
DB 2 RIAISGGTG-GHIYPALAFIKEVQ---RHPNVEFLYIGTENGLEKKIVVERENIPFSI 57
QY 52 EVTOYLVLKFSIGKQFYRLFYGVVDKIYNKRKFENIYFKMGNKRLGELVDEHOPDIINT 111
DB 58 EITGFKRLKSFENVKTVNR-----FLKGVKSKSYLAEPKDAVIGT 99
QY 112 FPMIVPPEYRRRTGRVLTFTNVTDFLHKIWHEN-----VDKYYVADYVKEKL-- 163
DB 100 GGYVCGPVVYAAAKMGIPTI-----VHEQNSLPGITNKF--LSKYVNVKVAICF 145
QY 164 -EIGTH-PS-NVITGPIRQPQESMPVPIYKYNLSPNKKVLLIMAGAHG---VLKN 217
DB 146 EEAKSHFPSEKVVFTGNP-RASEVVSITKGRSLAEFKLSEDKTKVLIFGSGRAAPINRA 204
QY 218 VKELCNLVKDDQVQVVVCGKNLTKESLSALEAEN-GDKLVGVYVERIDELFRITDC 276
DB 205 VIDM-QDVLTQVLYITGE-VHYEKVNLKSKGAADNVTKPLHOMPEYLKADV 262
QY 277 MITRPGGITTEAIGPVILYKVPV-----GOEKENAFFEDRGAAIYVNR-----HEEI 328
DB 263 IVARAGAAITAEITAGLIPSVLI-PSPYVTANHOEVNARSILGQHDAAIVLKETELSGEKL 321
QY 329 LESVTSLLADEDTLHRMKNKIKDLHLANSSE---VILEDILK 367
DB 322 IEALDRIVLNEQTLKEMSEKTSKLGVPDAAARLYSVLEELKK 363

RESULT 3
MURG_AQUAE STANDARD; PRT; 344 AA.
AC O67238;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE UDP-N-ACETYLGUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGUCOSAMINE TRANSFERASE
DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGCLNAC GLCNAC TRANSFERASE).
GN MURG OR AQ_1177.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC

MURG_BORBU		STANDARD;	PRT; 363 AA.
ID	MURG_BORBU		
AC	O51708;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	UDP-N-ACETYLGUCOSAMINE--N-ACETYLGLUCOSAMINE TRANSFERASE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGUCOSAMINE TRANSFERASE (EC 2.4.1.-) (UNDECAPRENYL-PYR-MURNAC-PENTAPEPTIDE-UDPGCLCNA GLCNAC TRANSFERASE).		
GN	MURG OR BB0767.		
GS	Borrelia burgdorferi (Lyme disease spirochete).		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.		
OX	NBCTaxID=139;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 35210 / B31;		
RX	MEDLINE=98065943; PubMed=9403685;		
RA	Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M., Dougherty J., Tomb J.-F., Fleischmann R.D., Richardson D., Peterson J., Karlavage A.R., Quackenbush J., Salzberg S., Hanson M., van Vugt R., Palmer N.L., Adams M.D., Gocayne J.D., Weidman J., Utterback T., Watthey L., McDonald P., Artiach P., Bowman C., Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B., Smith H.O., Venter J.C. ;		
RT	"Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi".		
RL	Nature 390:580-586(1997).		
-I-	SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MORNAC-(PENTAPEPTIDE) GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY). -I- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS. -I- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY). -I- SIMILARITY: BELONGS TO THE MURG FAMILY.		
-I-	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).		
EMBL; AE001176; AAC67113.1 ; -	TIGR; BB0767 ; -		
DR	Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane; Peptidoglycan synthesis.		
KW	Sequence 363 AA; 41118 MW; IBFA3437384DB235 CRC64;		
SQ			
Query Match	8.7%; Score 170.5; DB 1; Length 363;		
Best Local Similarity	20.6%; Pred. No. 7.7e-05;		
Matches	82; Conservative		
	Gaps 61; Indels 171; Mismatches 171; Gaps 16;		
QY	3 TNKRVLITANYNGHV-----OVAKLYEQCVLRGFQHVTYSNLVQENPIVSQTQLY 58		
Db	: :: : : :: : :		
Dy	2 SNKKIIFFTGGTGGTVFGPSISIQKLKEFNFEIEFFWGKNSTEER --LIKQDNTKF 59		
QY	59 LKSFSIGQFYRLFYYGVVDKYNNRKFNLYFK--MGNRKGELVDEHQPDIINFPDIV 116		
Db	: : : : : : : : : : : : : : : :		
Dy	60 I-SIPCGR-----LRRIFSANKTFDFKVLIIGSKFYKFQQIYATGGFVS 109		
QY	117 VPVEPRRGTRVIPTNVMTDFCLHKIWHE-----NDVKYIVATDYKVELLE-IG 166		
Db	: : : : : : : : : : : : : : : :		
Dy	110 TPAL-----IASLLKIKSIETHMDLPDLATKINSKPANNIHFSKESRYF 157		
QY	167 THPSNVKITGPIRPOFEEPMVGPIYKYNLSPNKVVILLIMAGAHG--VLKNVKRELCE 224		
Db	: : : : : : : : : : : : : : : :		
Dy	158 KNKNIILIYTSPRIREFLNPD--KIILQLTONTNKPRIISLUIGSGNALNNL-AIC-- 212		
QY	225 LVKDQDVQVWVCGKNTALKSELSALENBGDKLVGYVERIDFLFRITCDMITPKGI 284		

Query Match	8.7%	Score	171;	DB	1;	Length	351;
Best Local Similarity	23.9%	Pred. No.	6.8e-05;				
Matches	95;	Conservative	72;	Mismatches	150;	Indels	80;
Gaps	22;						

Qy	4	NKRVLILTANYGNHGV----	QVAKTLYEQ----	CVRLGFQHVTVSNLYQESNPVISEVTQY	56
Db	3	NKKLLVWAGGTG--GHVFPAINAQTLOQ	QEWDCWLG-----	TKDMEAAQLVPKVGIPIRF	57
Qy	57	LYLSFSIGKOFYRLFYGVNDKIYKRNKFN	IFYKMGKNKRLGELVDEHQPDIIINTFPMIV	116	
Db	58	IQISGLR--GK-----	GIKALLN--APFAIFRAVLOAK--	KIIQEEKPDAVLG-----	99
Qy	117	VPEYRRRTGRVIPFNVTDFCLHKIWHV	ENVDKYYVATDVVKKELL--EIGT-----	HPS	170
Db	100	-----MGYYVSGPAGVAAKLCGP	PIILHEQ-----	NAIAGLTNKLKGKATCATCVLQAPPT	148
Qy	171	--NVKINGIPIRPOFEESMPVGPIYKYN	LSPNKKVLLIMAGA-----	HGVLKNVKE	220
Db	149	APPHAEVGVNVPREDLFE--MENPDI--	RFSDBEEKLRLVVLVGGSGARVNLH	LPKVVAAQ	205
Qy	221	LCENLVKDDQVQVVVVGKNTALKE	SLSALBAENGDKLVLYGVERIDELPRIDCMITK	280	
Db	206	LADKLEFRHQV-----	KGGAV--EEVSQLYGENLEQVKITEFID	NNAAEAWADWVJCR	257
Qy	281	PGGITLTATATGVPVILYKVPVGOEKE	-----NANFFEDRGAATVVRNHE----	EILESV	332
Db	258	SGALTVCIBAAVGAIAIF--	VPEFHQRQVYLNKALYLSDDVGAAKII	EQADLTPEILVNY	314
Qy	333	TSLLADEDTLHRMKNKNIKDLHANS	SEVILEDKES	369	
Db	315	LKNLTRENLL--QMAKAKTMSMPNAAQ	RVAEVIKQYS	350	

Result	6
--------	---

```

Db 213 IKKDAEYFTHOSGKN-----LNDLSEKNVLRQFFN-AEEMASIVKFSNLIISRAG 265
QY 285 TLTEATAIGVPVILY---KVPQOEKENANFFEDRGAAIVVNRHE-----EILSEVSTLLA 337
Db 266 AIREFANAGACALIPPKKGRSGDQIKNAKLLTQNAQIYIDEDEILNINILKIKRTLK 325
QY 338 DEOTLHRMKNKIKDLHLANSSEVLEIDKESEMMTAK 375
Db 326 DREKINSKENIKFNKHSSTLIKLLIKDIKETS 363

RESULT 7
MURG_ENTHR
ID MURG_ENTHR STANDARD; PRT; 360 AA.
AC 007670;
DT 30-MAY-2000 (Rel. 39, Created)
DE 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLURAMYL-(PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE TRANSFERASE).
GN MURG.
OS Enterococcus hirae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SI85;
RX MEDLINE=99449055; PubMed=10520745;
RA Duez C., Thamm I., Sapunaric F., Coyette J., Ghuyssen J.-M.;
RT "The division and cell wall gene cluster of Enterococcus hirae SI85.";
RL DNA Seq. 9:149-161(1998).
CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y13922; CAAY4235.1;
CC KW Transferase; Glycosyltransferase; Cell division; Cell wall;
CC Membrane; Peptidoglycan synthesis.
CC SEQUENCE 360 AA; 39591 MW; 444D6F7CFEC1FDA1 CRC64;

Query Match 8.2%; Score 161.5; DB 1; Length 360;
Best Local Similarity 20.6%; Pred. No. 0.00032;
Matches 84; Conservative 69; Mismatches 149; Indels 105; Gaps 17;

QY 8 LIITANYGNHVOVAKTLYEQCVRLGFQHVTVSNLXQESNPVSEVTVYLYLKSF----- 62
Db 3 ILVTGGTGTHIYPA-----LAFVNVYKT---KEPN-----TEFMYVGAQRGLEN 44
QY 63 ----SIGKQYRLFYFYGVDK---LYNKRKFNIYFKMGNKELGELVDEHOPDIIINFPMI 115
Db 45 KIVPETGMPRTLEIOGFQRLKSLHNKLTQLFLK-SIREAKKILKEFKPDVWIGTGGY 103
QY 116 VPEYRRRTGRVPTFNVMVDFCLHIWHEH-----IHEQNSVPCITNKLRSRYVDRLTALSFDAAPF 151
Db 104 SGAVVYAASKLAIPTI-----IHEQNSVPCITNKLRSRYVDRLTALSFDAAPF 151
```

```

QY 157 YVKEKLEIGTHPSNVKITGIPRPOFEESMPVGPYKYKYNLSPNKKVLLIMAGAHGVK 216
Db 152 FPAEKSSLG-NP-----RAQEVADMDKSKILATYGLDPEKRTVLIFGSGOALK 200
QY 217 NVKELCENLVKDDQOVVVVCGKNTALKESLSALEAENGDKLVKLVGVYVERIDELFRITDC 276
Db 201 INQAVTEFLMSFDQEVLYASGERVYKDIQTKVPA--CANVSIQPIYNKMAEVMASSDL 258
QY 277 MITKPGGITLATEATGVPVILYKVPQOEKENANFFEDRGAAIVVNRHEEILSEVSTLL 336
Db 259 LVGRAGATSAELTALGLPAILI-PSP-----YVTNDHO--TKNAMSIV 299
QY 337 ADEDTLHRMKNKIKDLHLANS-----EVILEDILKESEMMTAK 377
Db 300 K-----NNAAKMIKDDLDORSLSKQAEIEINTNDQLOKQMSLASKQ 341

RESULT 8
MURG_STRCU
ID MURG_STRCU STANDARD; PRT; 362 AA.
AC Q9X4H4;
DT 30-MAY-2000 (Rel. 39, Created)
DE 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLURAMYL-(PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE TRANSFERASE).
GN MURG.
OS Streptomyces collinus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=42684;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 40733;
RA Mikulik K., Zhulanova E.;
RT "Characterization of dcw cluster from Streptomyces.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF110367; AAD26629.1;
CC KW Transferase; Glycosyltransferase; Cell division; Cell wall;
CC Membrane; Peptidoglycan synthesis.
CC SEQUENCE 362 AA; 38847 MW; 0BBA131F2E5D5122 CRC64;

Query Match 7.0%; Score 138.5; DB 1; Length 362;
Best Local Similarity 22.1%; Pred. No. 0.012;
Matches 46; Conservative 47; Mismatches 88; Indels 27; Gaps 8;

QY 168 HPSNVKITGI---PIRPOFEESMPVGPYKYKYNLSPNKKVLLIMAGAHGV--LKNVKELC 222
Db 158 HPVRRSIATLDRAAVPEAR-----ARFGLDPTLLVSGSGQARRLNEVYQV 208
QY 223 ENLVKDDQOVVVVCGKNTALKESLSALEAENGDKLV-LGVYVERIDELFRITDCMITKP 281
Db 209 APWLOQAGTQILHAVGP----KNELPQVQOMPQPIPVSYLDRMDLAYAADMMLCRA 264
```

QY	282	GGITLTETAIGVPVILYPVP--GQEKENANFFEDRGAIVVN-----RHEIILESVTS	334
Dt		: : : : : : : : : : : : : : :	
Db	265	GAMTVAELSAVGLP-AAYPLPTNGEQRNLNAPVVYKAGGLLVDDAELTPEWVOQTVLTP	323
QY	335	LLEADOTLHRMKNIKDLHLANSSEVIL	362
Dt		: : : : : : : : : : : : : : :	
Db	324	VLDAPHRLYEMSRAAGEFGRDRDAELLV	351
RESULT 9			
MURG_STRCO	ID	MURG_STRCO	STANDARD; PRT; 364 AA.
AC	QCZBA5;		
DT	30-MAY-2000	(Rel. 39, Created)	
DT	30-MAY-2000	(Rel. 39, Last sequence update)	
DT	01-OCT-2000	(Rel. 40, Last annotation update)	
DE	UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURMYL-(PENTAPEPTIDE)		
DE	PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE		
DE	(EC 2.4.1.-) (UNDECAPRENYL-P-P-MURNAC-PENTAPEPTIDE-UDPGLCNNAC GLCNAC		
DE	TRANSFERASE).		
GS	MURG OR SC4A10.17C.		
ON	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.		
OX	NCBI_TaxID=1902;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Kuennen R.A., Stadelmaier B.T., McCormick J.R.;		
RL	Submitted (JAN-1999), to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Saunders D.C.; Harris D., James K.D., Parkhill J., Bartell B.G.,		
RA	Rajadream M.A.;		
CC	Submitted (AUG-1999), to the EMBL/GenBank/DDBJ databases.		
CC	-1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC-		
CC	SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID		
CC	INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-		
CC	(PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).		
CC	-1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.		
CC	-1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE MURG FAMILY.		
CC	----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation		
CC	the European Bioinformatics Institute. There are no restrictions on re-use		
CC	use by non-profit institutions as long as its content is in accordance		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement		
CC	or send an email to license@isb-sib.ch).		
CC	----		
DR	EMBL; U10879; AADI0537.1; --		
DR	EMBL; AL109663; CAB51993.1; --		
KW	Transferase; Glycosyltransferase; Cell division; Cell wall;		
SW	Membrane; Peptidoglycan synthesis.		
QQ	SEQUENCE 364 AA; 38842 MW; D600C5F94E9202E2 CRC64;		

[illegible]

181 GLDPNPLTLLVTSGSQA--RLNEVIQQVAPWLQQAGIQILHAVGP-----KNELPQVHQM 235

254 NGOKLV-IGYVERIDELFRITCMITKGGITLTTAATGIVPVILYKVPV---GOEKEN 309

236 PGMPPIPVSYLDMDLAYAAAADMILKAGAMTVAELUSAVGLPA-AYVPLPIGNGEQLN 294

310 ANFFEDRGAAIVYN----RHEETLESVTSLLADEDTLHRMKNKIKOLHLANSSEVIL 362

295 AOPVWAGGSLLYDDAELTPEWLOQNVLPVLAOPHRLYEMSRAAAFGRRDADLLV 351

```

RESULT 10
MURG_MYCLE
ID MURG_MYCLE STANDARD; PRT; 407 AA.
AC 069552;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE-N-ACETYLURAMYL-(PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE (BC 2.4.1.1-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE TRANSFERASE).
DE MURG OR MLCB268.02C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1769;
RX [1]
SEQUENCE FROM N.A.
RA Brown D., Churcher C.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RL
CC -1- FUNCTION: CELL WALL FORMATION CATALYZES THE TRANSFER OF A GLCNAC
CC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL022602; CAA18668.1; -.
DR Transferase; Glycosyltransferase; Cell division; Cell wall;
KW Membrane; Peptidoglycan synthesis.
KW SEQUENCE 407 AA; 42422 MW; B757B73C092C53F9 CRC64;
SC SEQUENCE

```

	Query Match	7.0%; Score 137.5; DB 1;	Length 407;	
	Best Local Similarity	22.7%; Pred. No. 0.017;		
	Matches 68; Conservative	49; Mismatches 112;	Indels 71; Gaps 15;	
Qy	98 ELVDHEQHDPDIITFPMLVWPEYRRRTG----	RVIPTFNVMTDFCLHKIWHENVKXY	152	
	: : : : : : : : : : :	: : : : :		
Db	117 EVEEAH-----VVVFGGYVALFAYLAARGIPRVRRIP-----	VVVEANARAG	161	
Qy	153 VATDY---VKEKLEI---GTHPSNVKITGIPI-----	RPQEESMPVGIVKYKNLS	199	
	: : : : : : : : : : :	: : : : :		
Db	162 IANRGVGTAEURLVSAPGSLRGVAEYVGVPITHATITLNRPALR----	ADAREKHFGFT	216	
Qy	200 PNKKVILLMAGAHHVL---KNVKELCENLRKDQQVVVVCGKNTALKESIALEAENG	D	256	
	: : : : : : : : : : : : : :	: : : : : :		
Db	217 DDARLLVFSGSQSAVNRAVAGAAEDLASG-VAVIHAYG----	LKNTLELRTPEYGE	271	
Qy	257 KLVK-LGVVERIDELFRITDCMITKPGGITLTTEATAIGVPVILYKPVP---	GQEKENANF	312	
	: : : : : : : : : : :	: : : : : :		
Db	272 DEBVAAPVLDMDTAAADAIVICSSGAMTVAEVSAVGLPAL-VYPFFIGNCEORINALP	F	330	

db 14 GNGHVYRTLSLANELRKFNVNEIFFTTKSEDDVIKKTEENGFKVIKCSDDNDILKNIKNI 73

Query Match					
6.8%; Score 134; DB 1; Length 353;					
Best Local Similarity 20.0%; Pred. No. 0.024;					
Matches 80; Conservative 71; Mismatches 152; Indels 98; Gaps					
QY	10	LTANYGNHGVQVAKTYEQCVRLGFQHVTVSNLYQ-----ESNPIVSEVTQYLILKSF 63			
DB	5	LTGGGTGGHLSIAKALAELEKGIEAIVLGYTGDKWFPENSEPLFSE--RYFFNTQGV 62			
QY	64	TGQPVR----LFYGVVDKIYNRKKNIFPKMGNKRGELDEHPDIIINTFPVIWVE 119			
DB	63	VNKSPFKIRSLF-----LQAARAF-----KAKEILKKHOITHTISVGGSAGPA 107			
QY	120	YRRTRGRVIPFNVTDFCLHKIWWHE-----NVDKYY-----VATDYV-KEKLLEIG 166			
DB	108	SFASLLNKIP-----LYIHQNAIKGSLNRYLSPKAAVFSSYAFKDK---G 151			
QY	167	THPSNVKITGIPIRQPFESMPGYKYKYNLSPNKKVLLIMAGHG-----VLKNVK 219			
DB	152	NH----VLTSYPVQNAF-----FDHARTTEIKHLIFLGSGOGAKAINEFALLNAP 198			
QY	220	ELCENLVKDDQVVVVCGKNTALKSLSALEAENG--DKLAVLCYVERIDELFRITDCM 277			
DB	199	KLTGOGIK-----ITHICGPN--YEQVRFFYQELGLLDKIELFAFHNNITEVMHRADLC 251			
QY	278	ITKPGGIITLTAATGVPVILYKPVP-----GOEKENANFEEDRGAAYVNVRHE---EIL 329			
DB	252	VSRGASSWELCANGLPTT-FIPYPASNHHQYVNVLEFEKENLCYVVPQNELLPKKLF 310			
QY	330	ESVTSILLADE-----TLRMKKNIKDLHLANSSEVILEDL 366			
DB	311	EVIKLNQKDDQGKNLTITSAKLQOKIAKDGAKTIIETIL 351			
 RESULT 14					
ID	MURG_MYCTU	STANDARD;	PRT;	410 AA.	
AC	O06224;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	UDP-N-ACETYLGUCOSAMINE-N-ACETYLMURAMYL-(PENTAPEPTIDE)				
DE	UDPHOSPHORYL-UNDECAPRENOL N-ACETYLGUCOSAMINE TRANSFERASE				
DE	(EC 2.4.1.-) UNDECAPRENYL-P-P-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC				
DE	TRANSFERASE).				
GN	MURG OR RV2153C OR MTCY270.15.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
ON	NCBI_TaxID=1773;				
OX	[1]				
R	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RC	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock R., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies K., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.				
RA	Taylor K.G., Whitehead S., Barrrell B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
CC	-1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC				
CC	SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPI				
CC	INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-				
CC	(PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).				
CC	-1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.				
CC	-1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO THE MURG FAMILY.				

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 295388; CAB08640.1; -
DR Tuberculin; RV2153C; -
KW Transfrase; Glycosyltransferase; Cell division; Cell wall;
KW Membrane; Peptidoglycan synthesis.
SQ SEQUENCE 410 AA; 41860 MW; 96902AFE356FC30B CRC64;

Query Match 6.5%; Score 128.5; DB 1; Length 410;
Best Local Similarity 19.9%; Pred. No. 0.071;
Matches 54; Conservative 55; Mismatches 124; Indels 39; Gaps 11;
QY 98 ELVDEHQPDIIINFPIMIVPEY-----RRRTGRVPTFNVTDFCLHKIWH 145
DB 121 DVLDDVDADVVGFGYVALPAYLAAGLPLPRRRRIPVVIHEANARAGLA-NRVGAH 179
QY 146 ENVDKYYVATDVVKEKLEIGHGHPNVTGIPRQ----FEESMPVGPYIKKYNLSPNK 202
DB 180 -----TADRVLSAVPDGSLR--RAEVGVVPVRASTAAALDRAVLRAEAAHFGFPDDA 229
QY 203 KYLLIMAGAHGVLKNVKELCNLVKDDQVQVVVCGKNTALKESLSAL--EAENGDKLV 260
DB 230 RVLVFGSGQAV-SUNRAVSAAD--LAAAGCVLHANGQNVLELRRRAQGGPPYYA 286
QY 261 LGYVERIDELFRITCMITKPGGIIITLTAIGVPVILYKVPV---GOEKENANFFEDRG 317
DB 287 VPYLDRELAYAAADLVICRAGMTVAEVSAGVLPAL-YVPLPIGNIGEOQLNALPWNAG 345
QY 318 AAIIVN-----RHEEILESVTSLADETLHRM 345
DB 346 GGMVYADAALTPELVARQVAGLLTDPARLAAM 377

RESULT 15
NIFE_ANASP
ID NIFE_ANASP STANDARD; PRT; 480 AA.
AC Q41144;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFE.
GN NIFE.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RA Bulkema W.J., Scappino L.A., Haselkorn R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).
CC -!- PATHWAY: FE-MO COFACTOR BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFEN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U47055; AAA87947.1; -
DR InterPro; IPR000318; -
DR InterPro; IPR000510; -

DR pfam; PF00148; oxidored_nitro; 1.
DR PROSITE; PS00699; NITROGENASE_1_1; 1.
DR PROSITE; PS00090; NITROGENASE_1_2; 1.
KW Nitrogen fixation.
SQ SEQUENCE 480 AA; 52804 MW; F4F1D1BC968665DF CRC64;

Query Match 6.3%; Score 123.5; DB 1; Length 480;
Best Local Similarity 21.9%; Pred. No. 0.19;
Matches 60; Conservative 38; Mismatches 101; Indels 75; Gaps 9;
QY 73 YGVDKIYNKRKNFYFKMGNKRLGELVDEHQPDIIINFPIMIVVPEYRRRTGRVPTFN 132
DB 272 FYGIDEM-NRCLRNIAAKLGD-----PDLQARTEKLIARET 306
QY 133 VMTDFCLHKIWHNVKYYVATDVVKEKLEIGHGHPNVTGIPRQFEESMPVGP 192
DB 307 AALDLALAPYRARLKGKRVVLYTGGVKSWSIIISAAKDLGIEVYVATSTRKSTEE----- 359
QY 193 YKKYNLSPNKKVLLIMAGAHGVL---KNVKELCNLVKDDQVQVVVCGKN--TALKESL 247
DB 360 -----DKAKIKRLGADGIMLEKGNAKELLO-LVKDTQADMLIAGRNQYTALKARI 410
QY 248 SALEAENGDKLVGVERID---ELFRITDCMITKPGGIIITLTAIGVPVILYKVPV 304
DB 411 PFLDINQERHHPYAGYVGMIEARELYE-----ALYSPIWE 446
QY 305 QEKENANFFEDRG--AAIVNRRHEEILESVTSL 336
DB 447 QIRKPAPWDEDMGILAAHEYTSNHDHILASIEELI 480

Search completed: June 29, 2001, 09:04:10
Job time: 302 sec

7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 10:22:58 ; Search time 162.85 Seconds
(without alignments)
4430.205 Million cell updates/sec

Title: US-09-668-788-1

Perfect score: 1149

Sequence: 1 ttgaataccaataaaagag.....aagccaagtgtatcgtaa 1149

Scoring table:

IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601:*

1:	/SID88/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SID88/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SID88/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SID88/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SID88/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SID88/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SID88/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SID88/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SID88/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SID88/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SID88/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SID88/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SID88/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SID88/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SID88/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SID88/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SID88/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SID88/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SID88/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SID88/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	100.0	1149	20	AAZ23386
2	120.8	10.5	936	22	AAF58252
3	120.8	10.5	936	22	AAF58254
4	120.8	10.5	936	22	AAF58257
5	120.8	10.5	936	22	AAF58259
6	120.8	10.5	936	22	AAF58262
7	120.8	10.5	936	22	AAF58255
8	120	10.4	936	22	AAF58252
9	120	10.4	936	22	AAF58254
10	120	10.4	936	22	AAF58257
11	120	10.4	936	22	AAF58259

C 12	120	10.4	936	22	AAF58262	Oligonucleotide D2
C 13	120	10.4	936	22	AAF58255	Oligonucleotide D1
C 14	119.6	10.4	12658	18	AAV74438	Staphylococcus aur
C 15	54.2	4.7	244	22	AAF58238	Oligonucleotide D1
C 16	53	4.6	244	22	AAF58238	Oligonucleotide D1
C 17	48	4.2	1266	19	AAV12734	Monogalactosyldiac
C 18	41.6	3.6	1676	19	AAV43014	Streptococcus pneu
C 19	41.4	3.6	695	20	AAZ20376	Borrelia burgdorfe
C 20	41	3.6	3515	20	AAZ20252	Borrelia burgdorfe
C 21	40.4	3.5	5998	20	AAZ13056	Enterococcus faeca
C 22	39.8	3.5	10607	14	AAQ37894	Beta-casein coding
C 23	39.8	3.5	10607	19	AAV25594	Human beta-casein
C 24	39.6	3.4	452	18	AAV74893	Staphylococcus aur
C 25	39	3.4	4041	21	AAV70170	Plasmodium falcipa
C 26	38.8	3.4	1311	18	AAV74805	Staphylococcus aur
C 27	38.4	3.3	6532	21	AA81502	N. meningitidis pa
C 28	38.4	3.3	349980	21	AAZ21544	Neisseria meningit
C 29	38.2	3.3	2600	20	AAZ08512	MurE gene, Staphy
C 30	37.8	3.3	296	21	AAZ94179	Cat flea head and
C 31	37.4	3.3	861	20	AAZ99566	Nucleic acid seque
C 32	37	3.2	2849	20	AAZ20293	Borrelia burgdorfe
C 33	36.8	3.2	1764	21	AAZ47190	Arabidopsis thalia
C 34	36.8	3.2	2982	17	AAZ12999	Nitrate reductase
C 35	36.6	3.2	1410	20	AAZ61755	B. burgdorferi ant
C 36	36.6	3.2	910715	20	AAZ20248	Borrelia burgdorfe
C 37	36.2	3.2	1029	20	AAZ13600	Enterococcus faeca
C 38	36	3.1	2128	20	AAZ76373	P. carinii lysyl-t
C 39	35.6	3.1	909	22	AAZ88872	Human zcrp7 degen
C 40	35.4	3.1	2740	22	AAZ32681	Human cDNA encodin
C 41	35.4	3.1	3019	22	AAZ62010	Tomato LeSNF4 (suc
C 42	35.4	3.1	9062	18	AAZ74396	Staphylococcus aur
C 43	35.2	3.1	111309	20	AAZ20250	Borrelia burgdorfe
C 44	35	3.0	282	21	AAZ97854	Human EST clone DN
C 45	35	3.0	3771	21	AAA10741	C. albicans gene C

ALIGNMENTS

RESULT
AAZ23386
ID AAZ23386 standard; DNA; 1149 BP.
XX AAZ23386;
AC AAZ23386;

DT 13-DEC-1999 (first entry)

XX B. subtilis; glycosyl transferase catalytic domain DNA.

XX Catalytic domain; glycosyl transferase; processive activity; detergent;
KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
KW food emulsifier; polymer modifier; glycosyl transferase; ss.

XX Bacillus subtilis.

XX DE19819958-Al.

PD 30-SEP-1999.

XX 05-MAY-1998; 98DE-1019958.

XX 25-MAR-1998; 98DE-1013017.

XX (GVSE-) GVS; GES ERWERB & VERW LANDWIRTSCHAFTLICH.
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.

XX Wolter FP, Jorasz P, Heinz E, Zaehlinger U;

DR WPI; 1999-552364/47.

XX P-PSDB; AAY33444.

XX New glycosyl transferase protein, useful for producing glycolipids

PS Example 1; Page 6; 8pp; German.

xx This invention describes a novel protein (I) comprising identical or
cc different catalytic domains of glycosyl transferases and which has
cc processive activity (i.e. builds up oligosaccharide glycolipids by
cc successive addition of glucose units). (I) can be used to produce
cc 3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
cc modifiers or detergents. This sequence encodes a Bacillus subtilis
cc glycosyl transferase catalytic domain which is used in the method of
cc the invention.

xx Sequence 1149 BP; 380 A; 217 C; 244 G; 308 T; 0 other;

Query Match 100.0%; Score 1149; DB 20; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgaataccaataaaagagtagtaattttgactgcaaaattacggaatggacatgtgcag 60
Db 1 ttgaataccaataaaagagtagtaattttgactgcaaaattacggaatggacatgtgcag 60
Qy 61 gtagcccaaaacatttatgaacaatgtgtacggtcgttcgacgtataacagtttct 120
Db 61 gtagcccaaaacatttatgaacaatgtgtacggtcgttcgacgtataacagtttct 120
Qy 121 aatttgtaccagagcaaatccgattgtttcagaggttaactcaatcatttttaaaa 180
Db 121 aatttgtaccagagcaaatccgattgtttcagaggttaactcaatcatttttaaaa 180
Qy 181 agcttctcaatcgggaacagattttatcgtttgtttattacgaggttgacaaaatctat 240
Db 181 agcttctcaatcgggaacagattttatcgtttgtttattacgaggttgacaaaatctat 240
Qy 241 aataaacgtataatcaattatttactttaaaatgggtataaaagattggcggaactgtc 300
Db 241 aataaacgtataatcaattatttactttaaaatgggtataaaagattggcggaactgtc 300
Qy 301 gataacatcagccgagatattattataatcatttcgagtgatcgttcgccaataac 360
Db 301 gataacatcagccgagatattattataatcatttcgagtgatcgttcgccaataac 360
Qy 361 agacccgaactggaagagtagtctactccttcacgttcaacgttatgactgtttgtcttcat 420
Db 361 agacccgaactggaagagtagtctactccttcacgttcaacgttatgactgtttgtcttcat 420
Qy 421 aaaaatttggttcacgaaacgtgataaaattattatgtggcgacagattacgtgaaagaa 480
Db 421 aaaaatttggttcacgaaacgtgataaaattattatgtggcgacagattacgtgaaagaa 480
Qy 481 aaactgctggagatcggcactcatcccaagcaatgtataaaatcacaggaattccaatcagg 540
Db 481 aaactgctggagatcggcactcatcccaagcaatgtataaaatcacaggaattccaatcagg 540
Qy 541 ccgcaatttggaagaatccatccgtgtgtgcccgcgatatataaaagtagcaattcttcacca 600
Db 541 ccgcaatttggaagaatccatccgtgtgtgcccgcgatatataaaagtagcaattcttcacca 600
Qy 601 aacaaaaaagtgtcttctgatcatgtgacgtgtcacggtgtattaaagaaacgtataaagag 660
Db 601 aacaaaaaagtgtcttctgatcatgtgacgtgtcacggtgtattaaagaaacgtataaagag 660
Qy 661 ctgtgcgaaaacctgtcgaagtagtaccgaagtgttcgtgtgcgggaaaaaat 720
Db 661 ctgtgcgaaaacctgtcgaagtagtaccgaagtgttcgtgtgcgggaaaaaat 720
Qy 721 acggttttaaaagaatctttgagtcgcttgaaagcggaataatgttgacaaattaaaagtt 780
Db 721 acggttttaaaagaatctttgagtcgcttgaaagcggaataatgttgacaaattaaaagtt 780
Qy 781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatattaccaag 840
Db 781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatattaccaag 840

Qy 841 cccgcgccgacttactttgacagaagccacagccattgagtgccctgctcattctacaaa 900
Db 841 cccgcgccgacttactttgacagaagccacagccattgagtgccctgctcattctacaaa 900
Qy 901 cccgtgctgcccaggaagaaagaaatgcaaaactctttgaagaccgaggagctgccatc 960
Db 901 cccgtgctgcccaggaagaaagaaatgcaaaactctttgaagaccgaggagctgccatc 960
Qy 961 gttgtgaacgctcatgaagagattctcgtcagtcagtcacttcccttcttgcagatgaagat 1020
Db 961 gttgtgaacgctcatgaagagattctcgtcagtcagtcacttcccttcttgcagatgaagat 1020
Qy 1021 acctgtcatgcatgaagaaacattaaagaccttcatttagcaaaactcctctgaagtg 1080
Db 1021 acctgtcatgcatgaagaaacattaaagaccttcatttagcaaaactcctctgaagtg 1080
Qy 1081 atttagagatatcctgaagggaatcagaaatgatgaccgccaaacaaagccaaagtg 1140
Db 1081 atttagagatatcctgaagggaatcagaaatgatgaccgccaaacaaagccaaagtg 1140
Qy 1141 ctatcgttaa 1149
Db 1141 ctatcgttaa 1149

RESULT 2

AAF58252

ID AAF58252 standard; DNA; 936 BP.

XX AAF58252;

XX AC

XX 24-APR-2001 (first entry)

XX DE

XX Oligonucleotide D1835.

XX KW Electron-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX OS Synthetic.

XX WO200107665-A2.

XX PD 01-FEB-2001.

XX PF 26-JUL-2000; 2000WO-US20476.

XX PR 26-JUL-1999; 99US-0145695.

XX PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 10.5%; Score 120.8; DB 22; Length 936;


```
Db 664 WWWWWW... 723
QY 721 acggttttaaaagatttttgatgcgttgaaagcggaataatggtgacaaataaaagt 780
Db 724 WWWWWW... 783
QY 781 ctgggc 786
Db 784 WWWGC 789

RESULT 8
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
AC AAF58252;
XX
XX
DT 24-APR-2001 (first entry)
DE
DE Oligonucleotide D1835.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
KW
XX Synthetic.
OS
XX WO200107665-A2.
PN
XX
PD 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-US20476.
PF
XX 26-JUL-1999; 99US-0145695.
PR
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX Umek RM;
PI
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
SQ

Query Match 10.4%; Score 120; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.5e-23;
Matches 6; Conservative 482; Mismatches 292; Indels 0; Gaps 0;

QY 1 ttgaataccaataaaagaggttaatttgactgcaaatcaggaatcggaatggtgcag 60
Db 782 WWWWWW... 723
QY 61 gtgacaaaacatttatgaacaatgtgtacgctcggtttcagcatgtaacagtttct 120
Db 722 WWWWWW... 663
QY 121 aattgtaccagagtcacaaatccgattgttcagaggtacactcaatcatttataaaa 180
Db 662 WWWWWW... 603
```

```
QY 181 agcttccaatcggaaacagtttttatcggtttttattacggaggttgacaaaatctat 240
Db 602 WWWWWW... 543
QY 241 aataaacgtataatcaatatttacttttaaaatgggtataaaagattggcggaactgtc 300
Db 542 WWWWWW... 483
QY 301 gatgaacatcagcccgatattattattacattccgatgcgtgcgtccggaatac 360
Db 482 WWWWWW... 423
QY 361 agacgcgaactggaagtcattcctcactcctcaacgttatgactgattttgtctcat 420
Db 422 WWWWWW... 363
QY 421 aaatttggttcacgaaacgtggtgataataattattgtggcgacagattacgtgaaagaa 480
Db 362 WWWWWW... 303
QY 481 aaactgtggagatcggcactcactcaccagcaaatgtaaaatacacaggaattcccaatcagg 540
Db 302 WWWWWW... 243
QY 541 ccgcaatttgaagaatccatgcctgttgcccgatataataaaagatacaatctttcacca 600
Db 242 WWWWWW... 183
QY 601 acaaaaaagtgtcttcgatcatgcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db 182 WWWWWW... 123
QY 661 ctgtgcgaacacctgtcgaagatgaccaagtgcacagtgtgtcgtgcgtgcgggaaaaat 720
Db 122 WWWWWW... 63
QY 721 acggttttaaaagatttttgatgcgttgaaagcggaataatggtgacaaattaaagt 780
Db 62 WWWWWW... 3
```

```
RESULT 9
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
XX AC AAF58254;
XX
XX DT 24-APR-2001 (first entry)
XX
XX DE Oligonucleotide D1875.
XX
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX
XX OS Synthetic.
XX
XX WO200107665-A2.
XX
XX PD 01-FEB-2001.
XX
XX PF 26-JUL-2000; 2000WO-US20476.
XX
XX PR 26-JUL-1999; 99US-0145695.
XX
XX PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX PI Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT
```


Db 422 || 363
Qy 421 aaatttggttcacgaacgtgataataattattatgtgacagattacgtgaagaa 480
Db 362 303
Qy 481 aaactgtgagatcgccactcatcccaagcaatgtataaaatcacaggaattcccaatcagg 540
Db 302 243
Qy 541 ccgcaattgaagaatccatcgtctgtgcccgcgatatataaaagtgacaatcttccacca 600
Db 242 183
Qy 601 aacaaaagtgctctgatcatcgagtgctcacggtgtattaaagaacgtataaagag 660
Db 182 123
Qy 661 ctgtcgaaacacctgtcaaggatgaccaaagtgcaagtagtgcgtgtgcgggaaaaat 720
Db 122 63
Qy 721 acggtttaaaagaatcttgagtcgcttgaaagcggaagaaatggtgacaaatataaagtt 780
Db 62 3

RESULT 11
AAF58259/C
ID AAF58259 standard; DNA; 936 BP.
XX AAF58259;
XX 24-APR-2001 (first entry)
DE Oligonucleotide D2004.
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS WO200107665-A2.
PN 01-FEB-2001.
XX 26-JUL-2000; 2000WO-US20476.
XX 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX UmeK RM;
XX WPI; 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX Example 6; Page 128; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 10.4%; Score 120; DB 22; Length 936;
Best Local Similarity 0.8%; pred. No. 1.5e-23;
Matches 6; Conservative 482; Mismatches 292; Indels 0; Gaps 0;
Qy 1 ttgaatcccaataaagagattatttttgaactgcaaatcagcaaatgacatgtgcag 60
Db 782 723
Qy 611 gtagcgaacacactttatgaacaatgtgacgctgcgcttcacgatgtaaacatttct 120
Db 722 663
Qy 121 aattgtaccacaagagtcacaaatcccatgtttcagaggttaactcaatcactttatttaaa 180
Db 662 603
Qy 181 agcttcacacggaacaggttttatcgttgtttattacagagtgacaaaatctat 240
Db 602 543
Qy 241 aataaaacgtaaattcaatatttactttataaatggtaaaagatggcgcaactgtc 300
Db 542 483
Qy 301 gatgaacatcagcccgatattattattataacatttcgatgacgtgcgtaacac 360
Db 482 423
Qy 361 agacgcgaactggaagagtcattcctacctcactcaacgttatgactgattttgtcttc 420
Db 422 363
Qy 421 aaatttggttcacgaacgtgataataattattatgtggcgacagattacgtgaagaa 480
Db 362 303
Qy 481 aaactgtgagatcgccactcatcccaagcaatgtataaaatcacaggaatcccaatcagg 540
Db 302 243
Qy 541 ccgcaattgaagaatccatcgtctgtggcccgatatataaaagtcacaaatcttccacca 600
Db 242 183
Qy 601 acaaaaaagtgctctgatcatcgaggtgctcacggtgtattaaagaacgtataaagag 660
Db 182 123
Qy 661 ctgtcgaaacacctgtcaaggatgaccaaagtgcaagtagtgcgtgcgggaaaaat 720
Db 122 63
Qy 721 acggtttaaaagaatcttgagtcgcttgaaagcggaagaaatggtgacaaatataaagtt 780
Db 62 3

RESULT 12
AAF58262/C
ID AAF58262 standard; DNA; 936 BP.
XX AAF58262;
AC AAF58262;
XX 24-APR-2001 (first entry)
DT Oligonucleotide D2007.
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.


```
QY 121 aattgtaccagagtcacaaatccgattgttcagaggttaactcaatcactcttatttaaaa 180
Db 662 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 603
QY 181 agttctcaatcgcggaaacagttttatcgttgtttattacggagtgacaaaatctat 240
Db 602 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 543
QY 241 aataaacgttaattcaatatttactttaaatgggttaataaagattggcggaactgtc 300
Db 542 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 483
QY 301 gatgaacatcagcccgatattattataacatttcgtagtcgtagcgggaatac 360
Db 482 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 423
QY 361 agacgcgaactggaagagtcattctcactcactcactcactcactcactcactcactc 420
Db 422 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 363
QY 421 aaatttgggttcacgaacgtggataaattattgtggtgacagattacgtgaagaa 480
Db 362 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 303
QY 481 aaactgctgagatcggaacatccatccaaagcaatgtataaatacaggaattccaatcag 540
Db 302 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 243
QY 541 ccgaatttgaagaatccatcgctgtgtggtggtggtggtggtggtggtggtggtggtg 600
Db 242 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 183
QY 601 aacaaaagctgtctgtatcagcaggtgctcaggtgtattataaagaacgtgaaagag 660
Db 182 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 123
QY 661 ctgtgcgaacacctgtcgaagatgaccagagtgcaagtagtctgtgctggtcggaataat 720
Db 122 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 63
QY 721 acggctttaaagaatctttagtgcgtgagcgaagcgaagcgaagcgaagcgaagcga 780
Db 62 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 3

RESULT 14
AAV74438/c
XX ID AAV74438 standard; DNA: 12658 BP.
XX AC AAV74438;
XX DT 16-MAR-1999 (first entry)
XX ST Staphylococcus aureus contig SEQ ID #127.
XX KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX KW skin infection; surgical wound infection; scalded skin syndrome;
XX KW toxic shock syndrome; ds.
XX OS Staphylococcus aureus.
XX FH Key
XX misc_feature 181..240
FT /tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 1981..2040
FT /tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3781..3840
FT /tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 5581..5640
FT /tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 7381..7440
FT /tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 9181..9240
FT /tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 10981..11040
FT /tag= g
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT EP786519-A2.
XX 30-JUL-1997.
XX 07-JAN-1997; 97EP-0100117.
XX 05-JAN-1996; 96US-0009861.
XX (HUM- ) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX Rosen CA;
XX WPI; 1997-374922/35.
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX stored on computer readable medium and used in the production of
XX anti-S.aureus vaccines
XX Claim 1; Page 692-699; 3271pp; English.
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer readable
XX medium, preferably selected from a floppy or hard disk, random access
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX the S.aureus DNA sequences allows putative functions to be assigned so
XX that protein-encoding or regulatory regions of commercial, therapeutic or
XX industrial importance can be obtained. Specifically, sequences which are
XX likely to encode antigens have been identified and these polypeptides can
XX be used in a vaccine composition against S.aureus infection. The
XX polypeptides can also be used in a kit for the immunodetection of
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX skin and surgical wound infections, scalded skin syndrome, toxic shock
XX syndrome, etc. Organisms transformed with the DNA sequences can be used
XX for recombinant production of the polypeptides. The new DNA sequences
XX (and their fragments) are useful as primers or probes for isolating
XX homologues of any of the S.aureus DNA sequences contained on the
XX computer readable medium.
```


7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 09:32:17 ; Search time 1964.27 Seconds
(without alignments)
5529.448 Million cell updates/sec

Title: US-09-668-788-1
Perfect score: 1149
Sequence: 1 ttgaataccaataaaaggt.....aagccaaagtgcatacgtaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est1:*
42: em_est2:*
43: em_est3:*
44: em_est4:*
45: em_est5:*
46: em_est6:*
47: em_est7:*
48: em_est8:*
49: em_est9:*
50: em_est10:*
51: em_est11:*
52: em_est12:*
53: em_est13:*
54: em_est14:*
55: em_est15:*
56: em_est16:*
57: em_est17:*
58: em_est18:*
59: em_est19:*
60: em_est20:*
61: em_est21:*
62: em_est22:*
63: em_est23:*
64: em_est24:*
65: em_est25:*
66: em_est26:*
67: em_est27:*
68: em_est28:*
69: em_est29:*
70: em_est30:*
71: em_est31:*
72: em_est32:*
73: em_est33:*
74: em_est34:*
75: em_est35:*
76: em_est36:*
77: em_est37:*
78: em_est38:*
79: em_est39:*
80: em_est40:*
81: em_est41:*
82: em_est42:*
83: em_est43:*
84: em_est44:*
85: em_est45:*
86: em_est46:*
87: em_est47:*
88: em_est48:*
89: em_est49:*
90: em_est50:*
91: em_est51:*
92: em_est52:*
93: em_est53:*
94: em_est54:*
95: em_est55:*
96: em_est56:*
97: em_est57:*
98: em_est58:*
99: em_est59:*
100: em_est60:*
101: em_est61:*
102: em_est62:*
103: em_est63:*
104: em_est64:*
105: em_est65:*
106: em_est66:*
107: em_est67:*
108: em_est68:*
109: em_est69:*
110: em_est70:*
111: em_est71:*
112: em_est72:*
113: em_est73:*
114: em_est74:*
115: em_est75:*
116: em_est76:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

SOURCE          Zea mays.
ORGANISM        Zea mays
REFERENCE       Spermatophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        clade; Panicoidae; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
TITLE          1 (bases 1 to 610)
JOURNAL        Walbot,V.
COMMENT        Maize ESTs from various cDNA libraries sequenced at Stanford
                University
                Contact: Walbot V
                Department of Biological Sciences
                Stanford University
                855 California Ave, Palo Alto, CA 94304, USA
                Tel: 650 723 2227
                Fax: 650 725 8221
                Email: walbot@stanford.edu
                Plate: 614045 row: D column: 09.
FEATURES       Location/Qualifiers
                source
                1..610
                /organism="Zea mays"
                /cultivar="W23"
                /db_xref="taxon:4577"
                /clone_lib="614 - root cDNA library from Walbot Lab"
                /tissue_type="root"
                /dev_stage="3-4 days old"
                /lab_host="XL0LR"
                /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
                EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                Lab (LM)"
                Lab (LM) 182 a 145 c 107 g 175 t 1 others
BASE COUNT    182 a 145 c 107 g 175 t 1 others
ORIGIN
Query Match      3.8%; Score 44.2; DB 104; Length 610;
Best Local Similarity 49.8%; Pred. No. 0.36;
Matches 112; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 771 attaaagttctggcgtatgtggcgccattgagcgtatttcggtacacagattgcat 830
Db 535 AGTTCAGGTGAAGGCTTTGTACAAAGATGGAAGATGTATGGGTGCTTGTGATTGTAT 476
QY 831 gattaccagccgcgcgcattactttcacagaagccacagccattgagtcgtgcat 890
Db 475 CATTACAAGCGCAGCACCTGGTACAAATTCAGAGCAATGATCCGTGCTTACCAATAT 416
QY 891 ttgtacaaaccgtgcctggccggaagaaagaaacttctttgaagaccgcgg 950
Db 415 TCTAAATGATTATATTGCTGGACAGGAAGCTGGCAATGTCCTTACGTTGTGAAATGG 356
QY 951 agctgcacgtgtggaaccgtcatgaagagattctcgagtcagt 995
Db 355 ATGTGGGAAGTTCTCGAAATCTCCAAACAGATTGCGGAAGATAGT 311

RESULT 7
LOCUS PT007G23U/c 438 bp DNA GSS 02-NOV-2000
DEFINITION Parametium tetraurelia sequence M07D12u of the end of plasmid
PT007G23, genomic survey sequence.
ACCESSION AL447205
VERSION AL447205.1 GI:11122266
KEYWORDS GSS.
SOURCE Parametium tetraurelia.
ORGANISM Parametium tetraurelia
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
Parametium.
1 (bases 1 to 438)
AUTHORS Keller,A.M. and Cohen,J.
TITLE An indexed genomic library for Parametium complementation cloning
JOURNAL J. Eukaryot. Microbiol. 47 (1), 1-6 (2000)
MEDLINE 20114709
REFERENCE 2 (bases 1 to 438)

```

```

AUTHORS Dessen,P., Zagulski,M., Gromadka,R., Plattner,H., Kissmehl,R.,
Meyer,E., Beremier,M., Schultz,J.E., Linder,J., Pearman,R.E.,
Kung,C., Forney,J., Satir,B., Van Houten,J.L., Keller,A.M.,
Froissard,M., Sperling,L. and Cohen,J.
TITLE Parametium genome survey: a pilot project
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 438)
AUTHORS Gromadka,R. and Zagulski,M.
TITLE Random sequencing of the Parametium macronuclear genome
JOURNAL Unpublished
REMARK Institute of Biochemistry and Biophysics, Polish Academy of
Sciences, Warsaw, Poland
REFERENCE 4 (bases 1 to 438)
AUTHORS Cohen,J. and Sperling,L.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2000) Parametium Genome Survey Project, Centre de
Genetique Moleculaire, Centre National de la Recherche
Scientifique, 91198 Gif-sur-Yvette, France. E-mail:
Sperling@cgm.cnrs-gif.fr
COMMENT The present survey of the Parametium tetraurelia macronuclear
genome consists of end sequences of a library of random 4-12 kb
fragments obtained by Sau3A partial digestion of macronuclear DNA
clones in the BamHI site of pBSIKS+. See [4].
Genes are predicted from matches to other sequences. For more
information about this sequence or the Parametium Project, see
http://caroli.vjf.inserm.fr/pt.
FEATURES       Location/Qualifiers
                source
                1..438
                /organism="Parametium tetraurelia"
                /macronuclear
                /strain="stock d4-2"
                /db_xref="taxon:5888"
BASE COUNT    168 a 44 c 54 g 172 t
ORIGIN
Query Match      3.8%; Score 43.6; DB 258; Length 438;
Best Local Similarity 52.8%; Pred. No. 0.48;
Matches 94; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 105 gcatgtacagttcttaattgtaccagagatcaatccgattgttcagaggtaactca 164
Db 403 GAATTTATTGATTCGCGATTTATATAAAGTGTAAACAATCTTTTAAATGATTCAAA 344
QY 165 atactttatttaaaagcttctcaatcggaagaaacagtttattcggtttttattacgg 224
Db 343 AGTCCTAGATTACCAATGATTCATTTAGAAAATGAACATTAATTCCTTAATTTAT 284
QY 225 agttgacaaaatctataataaacgtataattcaattactttcttaaaatgggtaataaa 282
Db 283 TTTAGATTAGTATATATTTAAACCAAAATTAATATAGATTTCACAACTAAATTA 226

RESULT 8
LOCUS CNS0039G 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

```



```
COMMENT      Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mammeter in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              pl and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES     Location/Qualifiers
    source    1..1101
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone_lib="RPCI-98"
               /clone="BACR08K10"
               /note="end : TET3"

BASE COUNT   201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match      3.8%; Score 43.4; DB 219; Length 1101;
Best Local Similarity 14.1%; Pred. No. 0.7;
Matches 67; Conservative 213; Mismatches 195; Indels 0; Gaps 0;

QY 11 ataaagagcattaatttgactgcaaatcaggaatgagacatgtgcaggtgacccaaa 70
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1043 WWWWWWWATWDTWKKWWWWATAAKTDTATWTTWRTAWRADWAGRDGRGKGRDRDAATDA 984

QY 71 cactttatgacaatgtgtcagctgcgcttcagcatgtaacagttcttaattgtacc 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 983 DGAGRRDGGRRKKDKDKKGGDDKKKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 924

QY 131 aagagtcaaatccgattgttcagagtgaaactcaactcatttatttaaaagcttctcaa 190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 923 RKADDDDGAGDKDDGGKGDADDDTGTGDKDDDKDDKDDWDDKAKGTWGDATWAAATDWW 864

QY 191 tcgggaacacgtttatcgtttgtttattacagaggttgacaaatctataataaacgta 250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 WNGWADADWTTWDAADDWADWDAAWAWKWDADAWAGARTADRRDWDGDRAGKRGARK 804

QY 251 aattcaatatcttcttaaaatggtaataaaagattggcggaacttgcgtgaaacatc 310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 803 RRDRKRADRRDAADRDADAATWTTTTRTDWKKWKTDTWTRWAADRTWDRDDDDDD 744

QY 311 agcccgatatattataacattccgatgatcgtcgccggaatacacagacgcgaa 370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 743 RDRAGTAGRWRTWKRWRRTWDRDADADDTARDRRRRGGDDGADAGKGTGRKRR 684

QY 371 ctggaagatcattcctcactcaacgttatgactaattttgtctctataaaattggg 430
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 RRRATWDRTDAAWAAWTTTDTDDWKKDRRRKRGARRRRRTTARAADWWTWAKW 624

QY 431 ttacgaaaaacgtggtaaatattatgtgcgcacagattacgtgaaaggaact 485
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 623 DWAKWDKTRADRWDRWAADTWDARKADRDWAKARAWRARRRARARADRW 569

RESULT 9
CNS0039Q/c      1101 bp      DNA      GSS      03-JUN-1999
LOCUS          Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION      BACR08M12 of RPCI-98 library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION      AL063931
VERSION        AL063931.1 GI:4941788
KEYWORDS       GSS.
SOURCE         fruit fly.

ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
              1 (bases 1 to 1101)
              Genoscope.
              Direct Submission
              Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mammeter in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              pl and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES     Location/Qualifiers
    source    1..1101
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone_lib="RPCI-98"
               /clone="BACR08M12"
               /note="end : TET3"

BASE COUNT   260 a 155 c 97 g 385 t 204 others
ORIGIN

Query Match      3.8%; Score 43.4; DB 219; Length 1101;
Best Local Similarity 35.7%; Pred. No. 0.7;
Matches 99; Conservative 43; Mismatches 135; Indels 0; Gaps 0;

QY 176 taaaagctctcaatcggaagaaacagttttatcgtttgtttattacaggtgacaaaa 235
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1026 DAKARAKAAAAAAATTTGAWAAAAAAWTTAKAGTKGKTTKAGRGRAAAAAAATGCTT 967

QY 236 tctataaagcgttaaatcttaatttacttcttctcaacgttcaacgttatgactgtttgtc 295
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 966 AAAAAAAATWTAWAAWATTTTTTTTAAARRRGKKAATAAAWCTTGRKTTTT 907

QY 296 ttgtcgatgaacatcgccgatattattataacattccgatcgatcgctgcgcgg 355
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 906 TTWAAARTAAAAAAWATTTTAAAAAAGKAAAGTTAAWTTAAATWTAWAAATKGDGG 847

QY 356 aatcacagacgcgaactggaagatcattcctcactcaacgttatgactgtttgtc 415
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 846 GRARGGGGKTTTTTAAWAAAAAAWTTTTTTTAAATTTTTTWWAAAAATTTTGAW 787

QY 416 ttcataaatttggttcacgaacgtggataaata 452
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 786 TTTTAAAAAATTTTTTTTAAAAAAGKAAAGTTAAWTTAAATWTAWAAATKGDGG 750

RESULT 10
CNS016LI/c      1101 bp      DNA      GSS      26-JUL-1999
LOCUS          Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION      BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION      AL106896
VERSION        AL106896.1 GI:5624374
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM      Plasmid Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```


/clone_lib="Entamoeba histolytica sheared DNA"
/notes=vector: pHO51; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of *E. histolytica*
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a
method for isolate identification. *Exp. Parasitol.*
77:450). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (making small insert libraries for
whole genome shotgun sequencing projects. In genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1999)."
113 c 108 q 279 t

Query Match	3.7%	Score 42.4;	DB 247;	Length 852;
Best Local Similarity	51.6%;	Pred. No. 1.2;		
Matches	97; Conservative	0; Mismatches	91; Indels	0; Gaps
QY	161	ctcaataccctttatttaaaagcttctcaatcgcgggaacacgitttatcggttggtttttatt	220	
Db	581	CTTAAATCTTTTTTAAAAACACAATTTATTTCAGAACATATTTTTTGTTGTTTTTTTT	640	
QY	221	acggagtgtgacaaaatctataataacgttaaattccaattacttttaaatgggtaata	280	
Db	641	AAATTTTAAAGAAATTTTAAAGAAAATTTTAAAGAACAAATTTATAAAAAAAAAAACACTTA	700	
QY	281	aagattgggcgaacttgtcgatgaacatccagccgatattattattaacacatttcga	340	
Db	701	AAAAAAAAGGAATATTAAAGAATATCAATGATTATCATCTCTTTCCCTATTATTTTGGGA	760	
QY	341	tgaatcgtc	348	
Db	761	CTATCACC	768	

RESULT	14
CNS007BE	DNA
LOCUS	GSS . 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR15H24 of RPCI-98 library from Drosophila melanogaster (fruit fly); genomic survey sequence.
ACCESSION	AL066953
VERSION	.1 GI:4945517
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1001)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr - Web] : www.genoscope.cns.fr]
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

Mon Jul 2 09:00:02 2001

us-09-668-788-1.rst

Page 11

7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 09:56:28 ; Search time 119.12 Seconds
(without alignments)
1786.972 Million cell updates/sec

Title: US-09-668-788-1
Sequence: 1 ttgataccaataaagagt.....aagccaaagtctatcgttaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2.6/ptodata/1/lna/5A_COMB.seq: *
2: /cgn2.6/ptodata/1/lna/5B_COMB.seq: *
3: /cgn2.6/ptodata/1/lna/6A_COMB.seq: *
4: /cgn2.6/ptodata/1/lna/6B_COMB.seq: *
5: /cgn2.6/ptodata/1/lna/PTUS_COMB.seq: *
6: /cgn2.6/ptodata/1/lna/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49	4.3	7218	1	US-08-232-463-14
C 2	39.8	3.5	10607	1	US-08-078-090-3
C 3	38.2	3.3	2800	3	US-08-988-251-1
C 4	38.2	3.3	2600	4	US-09-386-048-1
C 5	36	3.1	2128	2	US-08-415-593-39
C 6	34.6	3.0	486	1	US-08-828-511-1
C 7	34.4	3.0	1991	2	US-08-415-593-40
C 8	34.2	3.0	5852	1	US-07-867-106-2
C 9	34	3.0	1000	1	US-08-599-252-96
C 10	34	3.0	1000	5	PCR-US96-06332-96
C 11	34	3.0	1000	5	PCR-US96-06583-96
C 12	33.8	2.9	1395	1	US-07-991-867B-25
C 13	33.8	2.9	1395	2	US-08-107-755A-25
C 14	33.8	2.9	1395	2	US-08-544-332-25
C 15	33.8	2.9	6768	1	US-08-107-755A-1
C 16	33.8	2.9	8457	1	US-07-991-867B-1
C 17	33.8	2.9	8457	2	US-08-544-332-1
C 18	32.8	2.9	174	2	US-08-378-235B-8
C 19	32.8	2.9	296	4	US-09-235-614-38
C 20	32.8	2.9	615	4	US-08-998-416-186
C 21	32.8	2.9	1107	1	US-08-165-038-3
C 22	32.8	2.9	1107	1	US-08-621-081A-3
C 23	32.8	2.9	1107	2	US-08-876-781-3
C 24	32.8	2.9	1384	4	US-09-058-489-30
C 25	32.8	2.9	2945	6	5196526-2
C 26	32.8	2.9	7560	4	US-09-103-478-4
C 27	32.4	2.8	3350	4	US-08-617-860B-3

28	32.4	2.8	9048	4	US-08-973-273-4	Sequence 4, Appli
C 29	32.2	2.8	389	3	US-08-851-843A-62	Sequence 62, Appli
C 30	32.2	2.8	389	4	US-08-974-549A-8	Sequence 8, Appli
C 31	32.2	2.8	2171	3	US-08-851-843A-100	Sequence 100, App
C 32	32.2	2.8	2171	4	US-08-974-549A-266	Sequence 266, App
C 33	32.2	2.8	2176	4	US-08-974-549A-3	Sequence 3, Appli
C 34	32.2	2.8	3855	4	US-08-974-549A-4	Sequence 4, Appli
C 35	32.2	2.8	4015	3	US-08-851-843A-224	Sequence 224, App
C 36	32.2	2.8	4015	4	US-08-974-549A-1	Sequence 1, Appli
C 37	32.2	2.8	4029	3	US-08-851-843A-173	Sequence 173, App
C 38	32.2	2.8	4029	4	US-08-974-549A-292	Sequence 292, App
C 39	32.2	2.8	4037	4	US-08-974-549A-343	Sequence 343, App
C 40	32.2	2.8	6669	4	US-09-212-971-5	Sequence 5, Appli
C 41	32.2	2.8	6669	4	US-08-800-929A-5	Sequence 5, Appli
C 42	32	2.8	5496	1	US-08-181-629A-2	Sequence 2, Appli
C 43	31.8	2.8	2435	4	US-09-306-593-1	Sequence 1, Appli
C 44	31.6	2.8	7101	1	US-08-480-604A-9	Sequence 9, Appli
C 45	31.6	2.8	7101	2	US-08-405-496A-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-F1s
US-08-232-463-14

TELEFAX: 617-862-9540
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 2128 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-415-593-39

Query Match 3.1%; Score 36; DB 2; Length 2128;
Best Local Similarity 49.5%; Pred. No. 1.1;
Matches 93; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 146 tigtgttcagaggaactcaactcttttataaaagcttctcaatcgaggaaacagtttt 205
Db 2107 TTTTATATATTTTAAATACATACATAGATTATTTTACAAAGATTATATACTTTAAATAATACAC 2048
QY 206 atcggtttgtttattacggaggttgacaaactctataataaactgtaaaattcaatttact 265
Db 2047 TTTTATATATTTTAAATATTTTAAATATTTTATATATATATATATATATATATATATATAT 1988
QY 266 ttaaatgggtaataaagattggcggaacttgcgtgaacatcagcccgatattatta 325
Db 1987 TTTCAAAATATATGCAAAATTAAGAGTTATATTTTGAAATATATTTTCAAAATATTT 1928
QY 326 ttaataca 333
Db 1927 TTATTTAA 1920

RESULT 6
US-08-828-511-1
Sequence 1, Application US/08828511
Patent No. 5731193
GENERAL INFORMATION:
APPLICANT: MORI, Tetsuya
APPLICANT: YAMAMOTO, Kozo
APPLICANT: OHTA, Tsunetaka
TITLE OF INVENTION: RECOMBINANT DNA AND TRANSFORMANT
TITLE OF INVENTION: CONTAINING THE SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,511
FILING DATE: 31-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/355,245
FILING DATE: 09-DEC-1994
APPLICATION NUMBER: JP 342237/1993
FILING DATE: 15-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: MORI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-828-511-1

Query Match 3.0%; Score 34.6; DB 1; Length 486;
Best Local Similarity 51.6%; Pred. No. 1.5;
Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 138 aaatcgaggttttcagaggtaactcaactcttttataaaagcttctcaatcgaggaa 197
Db 20 AAAATCTATTTTCAAAAAGTTGCTTAAGAAATATAGTTATCAAGTTAAGTAAATGTCA 79
QY 198 acagttttatcggtttgtttttattacggaggttgacaaactctataataaactgtaaaattcaa 257
Db 80 ATAGCCTTTTAAATTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 139
QY 258 tattacttttaaatgggtaataaaagattggg 290
Db 140 TTTACTTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 172

RESULT 7
US-08-415-593-40/c
Sequence 40, Application US/08415593
Patent No. 5912140
Patent No. 5912140 5776726
GENERAL INFORMATION:
APPLICANT: Whoriskey, Susan K.
APPLICANT: Quinn, Cheryl L.
APPLICANT: Tao, Nijun
APPLICANT: Pollis-Virk, Karen I.
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA
TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,593
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI94-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-862-9540
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1991 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 29..1780
US-08-415-593-40

```
Query Match 3.0%; Score 34.4; DB 2; Length 1991;
Best Local Similarity 48.9%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 146 ttgttcagaggttaactcaatcaccttttttttaaaagcttctcaatcggaacagtttt 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1977 TTTTATTTTAAATACATACATAGATTATTTTACAGAAATTTTAACTTTAAATAATCAC 1918

QY 206 atcgttggttttattacggaggttgacaaaatctataataaacgtaaatcaattttact 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1917 TTTTAAATATATTTTAACTAAATTTTAACTTTTATAATAACAAATTTTCTTAAATTTT 1858

QY 266 ttaaaatgggtaataaagattggcgaaacttgctcgatgaacatcagcccgatatttta 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1957 TTTCAAAATATATGCAAAATAGAGTTATATTTTGAATAATATTTTCAAAATATTT 1798

QY 326 ttaataca 333
    ||| ||| |||
Db 1797 TTAATAAA 1790

RESULT 8
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; NAME/KEY: CDS
```

```
; LOCATION: 2378..5038
US-07-867-106-2

Query Match 3.0%; Score 34.2; DB 1; Length 5852;
Best Local Similarity 46.4%; Pred. No. 5.3;
Matches 111; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 200 agttttatcgtttgtttattacggaggttgacaaaatctataataaacgtaaatcaata 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1859 AATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1918

QY 260 ttactttaaaatgggtaataaagattggcgaaacttgctcgatgaacatcagcccgata 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1919 GTTATTTTAAATATGCAAAACTATAAAACTAATAGTAGTTTAACTTTTCTA 1978

QY 320 tttattataacatttccgatgatcgtgcccgggaataacagacgacgtaactggaagag 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1979 TTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2038

QY 380 tcattctacttcaacggttatgactgattttgtcttcataaaatttgggttcacgaa 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2039 AAAACCCCTCATTATAATAATTAATTACTTTGGTTTGTGATTTTATTTTATTTTAA 2097

RESULT 9
US-08-599-252-96/c
; Sequence 96, Application US/08599252
; Patent No. 5705343
; GENERAL INFORMATION:
; APPLICANT: DRYNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GINKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFF, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,252
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.21
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-599-252-96
```

Query Match 3.0%; Score 34; DB 1; Length 1000;

Best Local Similarity 49.4%; Pred. No. 2.9;
Matches 88; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 157 gtaactcaatacctttatttaaaagcttctcaatcggggaacagttttatcggtttgtt 216
Db 242 GTGACATTAATTTTAAATTAATAATATTTTATAATGTGACATATTATAAATATTATAA 183
QY 217 tattacggagttgacaaatctataaagcgtaaattcaatatttacttttaaaatgggt 276
Db 182 TAATAAATAATAGTAATAATAATATTTTAAATTAATATTTTAAATAATAAT 123
QY 277 aataaaagattggcgaaacttgcgatgaacatcgccgcgatattattattataacat 334
Db 122 AAATAAATAATGAATAATATGACATTAAATAATAAGAGTACTACTACTACTAATCCAT 65

RESULT 10

PCT-US96-06352-96/c

; Sequence 96, Application PC/TUS9606352

; GENERAL INFORMATION:

; APPLICANT: DRAYNA, DENNIS T.

; APPLICANT: FEDER, JOHN N.

; APPLICANT: GNIRKE, ANDREAS

; APPLICANT: KIMMEL, BRUCE E.

; APPLICANT: THOMAS, WINSTON J.

; APPLICANT: WOLFF, ROGER K.

; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY

; NUMBER OF SEQUENCES: 124

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20006-1888

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/06352

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/599,252

; FILING DATE: 09-FEB-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: MURASHIGE, KATE H.

; REGISTRATION NUMBER: 29,959

; REFERENCE/DOCKET NUMBER: 9053-0001.21

; TELEPHONE: (202) 887-1500

; TELEFAX: (202) 887-0763

; TELEX: 90-4030

; INFORMATION FOR SEQ ID NO: 96:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1000 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

PCT-US96-06352-96

Query Match 3.0%; Score 34; DB 5; Length 1000;

Best Local Similarity 49.4%; Pred. No. 2.9;

Matches 88; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 157 gtaactcaatacctttatttaaaagcttctcaatcggggaacagttttatcggtttgtt 216
Db 242 GTGACATTAATTTTAAATTAATAATATTTTATAATGTGACATATTATAAATATTATAA 183

Qy 324 tattaatacatticcgatgat 344
||||| | | | |
Db 572 TATTGAATCTGATATGACTAT 552

Search completed: June 29, 2001, 11:46:43
Job time: 6615 sec

7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:56:07 ; Search time 35.83 Seconds
(without alignments)
646.340 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILITANYNGHVQ.....EDILKESENMTAKQKAKVLS 382

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1970	100.0	382	20	AAY33444 B. subtilis glycosyl transferase
2	326	16.5	522	21	AA19039 Amino acid sequence
3	309.5	15.7	492	21	AAG42414 Arabidopsis thaliana
4	309.5	15.7	551	21	AAG42413 Arabidopsis thaliana
5	305.5	15.5	422	19	AA41148 Monogalactosylidase
6	303.5	15.4	525	21	AA19040 Amino acid sequence
7	301.5	15.3	533	21	AA19041 Amino acid sequence
8	293	14.9	468	21	AA19042 Amino acid sequence
9	287.5	14.6	404	21	AAG42415 Arabidopsis thaliana
10	114.5	5.8	468	21	AA92952 Arabidopsis thaliana
11	113.5	5.8	355	21	AA974401 Neisseria gonorrhoeae

12	112.5	5.7	993	19	AA98240 H. pylori GHPO 741
13	112.5	5.7	1068	13	AA20198 Sucrose phosphate
14	112.5	5.7	1068	18	AA38266 Maize sucrose phosphate
15	112.5	5.7	1068	18	AAW0869 Sucrose phosphate
16	111	5.6	478	20	AA92950 WO905287 Seq ID 1
17	110	5.6	469	21	AA92950 Arabidopsis thaliana
18	110	5.6	503	21	AA92950 Arabidopsis thaliana
19	110	5.6	519	21	AA92950 Arabidopsis thaliana
20	110	5.6	542	21	AA92950 Arabidopsis thaliana
21	110	5.6	576	21	AA92950 Arabidopsis thaliana
22	110	5.6	592	21	AA92950 Arabidopsis thaliana
23	109	5.5	609	13	AA92950 Arabidopsis thaliana
24	109	5.5	690	13	AA92950 Arabidopsis thaliana
25	109	5.5	754	13	AA92950 Arabidopsis thaliana
26	109	5.5	800	13	AA92950 Arabidopsis thaliana
27	109	5.5	856	13	AA92950 Arabidopsis thaliana
28	108.5	5.5	372	19	AA92950 Arabidopsis thaliana
29	108.5	5.5	1068	13	AA92950 Arabidopsis thaliana
30	108	5.5	892	13	AA92950 Arabidopsis thaliana
31	108	5.5	892	13	AA92950 Arabidopsis thaliana
32	107.5	5.5	1411	17	AA92950 Arabidopsis thaliana
33	107	5.4	552	20	AA92950 Arabidopsis thaliana
34	107	5.4	668	21	AA92950 Arabidopsis thaliana
35	106	5.4	1254	11	AA92950 Arabidopsis thaliana
36	106	5.4	1254	18	AA92950 Arabidopsis thaliana
37	104.5	5.3	473	19	AA92950 Arabidopsis thaliana
38	104.5	5.3	1198	19	AA92950 Arabidopsis thaliana
39	104	5.3	403	18	AA92950 Arabidopsis thaliana
40	104	5.3	1634	21	AA92950 Arabidopsis thaliana
41	104	5.3	1634	21	AA92950 Arabidopsis thaliana
42	102.5	5.2	367	21	AA92950 Arabidopsis thaliana
43	102.5	5.2	416	20	AA92950 Arabidopsis thaliana
44	102	5.2	1123	21	AA92950 Arabidopsis thaliana
45	101.5	5.2	340	21	AA92950 Arabidopsis thaliana

ALIGNMENTS

RESULT 1
AAY33444
ID AAY33444 standard; Protein; 382 AA.
XX
AC AAY33444;
XX
DT 13-DEC-1999 (first entry)
XX
DE B. subtilis glycosyl transferase catalytic domain.
XX
KW Catalytic domain; glycosyl transferase; processive activity; detergent;
KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
KW food emulsifier; polymer modifier; glycosyl transferase.
XX
OS Bacillus subtilis.
XX
PN DE19819958-A1.
XX
PD 30-SEP-1999.
XX
PF 05-MAY-1998; 98DE-1019958.
XX
PR 25-MAR-1998; 98DE-1013017.
XX
PA (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Wolter FP, Jorasch P, Heinz E, Zaehring U;
DR WPI; 1999-552364/47.
XX N-PSDB; AAZ23386.
XX
PT New glycosyl transferase protein, useful for producing glycolipids -
XX

PS Example 1: Page 7: 8pp; German.

This invention describes a novel protein (I) comprising identical or different catalytic domains of glycosyl transferases and which has successive activity (i.e. builds up oligosaccharide glycolipids by successive addition of glucose units). (I) can be used to produce 3-oligo-glucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer modifiers or detergents. This sequence represents a Bacillus subtilis glycosyl transferase catalytic domain which is used in the method of the invention.

Sequence	382 AA:
50	

Query Match	100.0%	Score 1970;	DB 20;	Length 382;
Best Local Similarity	100.0%;	Pred. No. 1.6e-180;		
Matches 382;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0;

Qy	1	mntnkrvllltanygnghvqvaktlyecovrlgfhqhtvsnlqesnpiysevtqzlylk	60
Db	1	mntnkrvllltanygnghvqvaktlyecvzrlgfhvtsvsnlqesnpiysevtqzlylk	60
Qy	61	sfsgikqofvrlfygyvdktyinkrkknfyfkmgnkrlgelvdehqpdiiintfpmivvpey	120
Db	61	sfsgikqofvrlfygyvdklynrkrknfyfkmgnkrlgelvdehqpdiiintfpmivvpey	120
Qy	121	rrrtgrvriptfnvmtdfclhktwvhenydkyyvafdyvkekllbeighpnsnkitgipir	180
Db	121	rrrtgrvriptfnvmtdfclhkiwhvhenydkyyvafdyvkekllbeighpnsnkitgipir	180
Qy	181	pqfeesmpvgpipykknynlspnkvlllmagahgvlknvkelcennlvkddovvvvvcgkn	240
Db	181	pqfeesmpvgpilykknynlspnkvlllmagahgvlknvkelcennlvkddqvvvvvcgkn	240
Qy	241	talkeslsaleaengoklvlgveridelfrtcdmktpggltlteaigaivpvilyk	300
Db	241	talkeslsaleaengoklvlgveridelfrtcdmktpggltlteaigaivpvilyk	300
Qy	301	pvpgoekenanffedrgaaivynrhrheilesvtslladedtlhrmkknikdlhlanssev	360
Db	301	pvpgoekenanffedrgaaivynrhrheilesvtslladedtlhrmkknikdlhlanssev	360
Qy	361	iledilkeseemmtakokavls	382
Db	361	iledilkeseemmtakokavls	382

RESULT

AAB19039
ID AAB19039 standard; protein; 522 AA.
XX
AC AAB19039:

08-FEB-2001 (first entry)

XX Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
DE
XX
XX Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW acquired parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW acquired immune deficiency syndrome; coccidiosis.
KW

XX
30
sninacia oleracea.XX
DNI
W0200056919-A1

XX 000-2000-SEP-29

XX
XX
XX

XX
10 1000
0000-000313A

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX

PT Marechal E, Block M, Joyard J, Douce R;

WPT: 2000-602227/57.

Use of monogalactosyl diacylglycerol synthase for identifying its specific inhibitors, potentially useful as antiparasitic agents and herbicides -

Example 2: Fig 2: 33pp; French.

The present sequence represents a monogalactosyl-diacylglycerol (MGDG) synthase. MGDG is present in all plasids tested and is essential for cell survival, but is not present in other membrane structures, or in animal cells, and so represents a specific target. MGDG synthase or a plastid membrane isolated from plants is used to select and screen for specific inhibitors of MGDG synthase. These inhibitors are suitable as active agents against apicomplex parasites or as herbicides. The inhibitors are used as antiparasitic agents, especially against Plasmodium, Toxoplasma (particularly in patients with acquired immune deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and as herbicides.

AA	Sequence	522 AA;
SQ		

Query Match 16.58: Score 326; DB 21; Length 522;

Query Match	10.38	Score: 1207	22	Length: 100
Best Local Similarity	25.3%	Pred. No.	1.2e-22	
Matches	98	Conservative	80	Mismatches 162;
Indels	48	Gaps	12;	

5 KRVI.TI.TANYGNHVOVAKTLYEOCVRLGFQ-----HVTVSNLYQESN--PIVSEVTQ 55

131 krylilmsdtgaghrasa-----eaiikaafneefaddvqvftdlwsehtowpfnqlprs 18

56 YIYIKSEICKOEYBIEVYGV-D-KTVNKRKFNTYFKMGKNRKGELVDEHOPDIIINTFPM 11

[illegible]

16

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051

```
QY      ZZ / RDDQA
|::||   |::|   :::||   |:|:|:|:|
|::||   |::|   :::||   |:|:|:|:|
```

[illegible]

QY 282 GGITL TEATAIGVPVILYKPPVQGERENANFFEDRGAAI VVNKHEEILEESVISEADEDI 3

Db 421 gpgtiaeamirglpiliindylagqaeagnvpvyiengigkyikspkeiaktvsqwirgpk--4

QY 342 LHRMKNIKDLHLANSSEVILEDILKES 369

Db 479 -----anelqimsqnaalkha 493

3 RESULT

RESULTS
AAG42414

ID
 YX
 AFAG42414

AC
yy
AAG42414;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52894.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;;
KW termination sequence.

OS *Arabidopsis thaliana*.

PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	28-OCT-1999;	99US-0162142.	
Query Match			15.7%; Score 309.5; DB 21; Length 492;
Best Local Similarity			24.2%; Pred. NO. 4.le-21;
Matches 102; Conservative			91; Mismatches 171; Indels 57; Gaps 17;
QY	5	KVLLILITANYGNHGVQAKTL-----YEQCVRLGFO---HVTVSNLQ 44	
DB	83	kkvllmsdtggghrasaeairaafnqefgdeyqsglefkigfrcvlhvftldlwt 142	
QY	45	ESN--PIVSEVTQYLYLKSFSIGKQFYRLFYGVDP-KIYNKRKFNIYFKMGNKRLGELVD 101	
DB	143	dhtpwpfnqlprsynflvkgh---tlwmttygtprlvghnsfaatstfiarelaqglm 199	
QY	102	EHQPDIIINTFPMI--VWPEYRRRTG--RVIPTFNMWDF-CLHKIWHNVNDKYVYVATD 156	
DB	200	kygpdliisvhpmlghvplrvlrskgllkivttvtdlstdtchptfwhklvtrcycpst 259	
QY	157	YVKEKLEIGTHPSNVKTINGIPRQFESMPVGP---IYKYNLSPNKKVLLIMAGAHG 213	
DB	260	evakraqkagletsikivylpvprpsfvk--pvprkvelrrlgmdenlpavilmgg9eg 317	
QY	214	V--LKNVKELCNLVKDDQV-----QVVVCGKNTALKESLSALEANGDKLKLGVYVER 266	
DB	318	mpieataraladalydnlgaeavqyliicgrnkklsldwk--ipvqvkfityk 375	
QY	267	IDELFRITDCMTKPGIITLTAIGVPIVLYKVPVGOEKENANFFEDRGAAIIVVNRHE 326	
DB	376	meecmgacdiitkagptiaeamirgipilingylagaeagnvpyvvengc----gkfs 431	
QY	327	EILESTVSLA0-----EDTLHRMKKNIKDLHLANSSEVILEDKLSEMMTKAKVYL 381	
DB	432	kspkelskiadvfgpaskelsqna--irla-kpeavfkivhdmlvtrkksnlpq 488	

QY	382 S 382	
DB	489 s 489	
RESULT 4		
AAG42413		
ID	AAG42413 standard; Protein; 551 AA.	
XX	AAG42413;	
AC		
XX	18-OCT-2000 (first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 52893.	
XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
OS		
XX	EP1033405-A2.	
PN		
XX	06-SEP-2000.	
PD		
XX	25-FEB-2000; 2000EP-0301439.	
PF		
XX	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.

Db	202	dhtpwpfnqlprsynflvkhg-----tlwkmtygtgsprivghnsfaatstfiarelaaglm	258
Qy	102	EHQPDIIINTFPMI--VVPYRRRTG--RVIPTFNMTDF--CLHKIWHVHNVDKYYVATD	156
Db	259	kyqpdliisvhlpmqhvplrvlrskgllkkkivftitdlstcptwfhkiivtrcypst	318
Qy	157	YVKEKLEIGHTHSPNVKTIQIPRPOFESMPVGP---IYKKYNLSPNKKVLLIMAGAHG	213
Db	319	evakraqkagletsqkvvglpvrpsfvk--pvrpkvelrrelgmdenlpavllimgggeg	376
Qy	214	V--LKNVKELCENLVKDDOV-----QVVVVCCKNTALKESLSALEAENGDKLVGLGYVER	266
Db	377	mgpieataraladalydknlgeavggylicgrnkkksgkssidwk--ipvqvkgrftk	434
Qy	267	IDELFRITDCMTYKPGGIIITRATATGVPVILYKVPVQGEKENANFFEDRGAALVVRHE	326
Db	435	meemgacdciiitkagpgtiaeamirglpilingiaggeagnpvpyvengc---gkfs	490
Qy	327	ELLESVTSLLAD-----EDTLHRMKKNIKDLHLANSSEVILEDLIKSEMMTAKOKAKVL	381
Db	491	kspkeiskivadvfgpaskelimsqna--lrla-kpeavfkivdhmhelvrkknslpq	547
Qy	382	s 382	
Db	548	s 548	
RESULT 5			
AAW41148			
ID	AAW41148	standard; Protein; 422 AA.	
XX	AC	AAW41148;	
XX	DT	05-MAY-1998 (first entry)	
XX	DE	Monogalactosyldiacylglycerol synthetase sequence.	
XX	KW	Monogalactosyldiacylglycerol synthetase; MGDG; lipid production;	
XX	KW	cucumber.	
XX	OS	Cucumis sativus.	
XX	PN	JP10014579-A.	
XX	PD	20-JAN-1998.	
XX	PF	02-JUL-1996; 96JP-0172337.	
XX	PR	02-JUL-1996; 96JP-0172337.	
XX	PA	(KIRI) KIRIN BREWERY KK.	
XX	DR	WPI; 1998-138241/13.	
XX	DR	N-PSDB; AAV12734.	
XX	PT	New mono:galactosyl-di:acyl-glycerol synthase - useful for, e.g.	
XX	PT	increasing lipid production in transformed host	
XX	PS	Claim 1; Page 8; 13pp; Japanese.	
XX	CC	This sequence represents the monogalactosyldiacylglycerol synthase (MGDG)	
XX	CC	of the invention. This sequence was isolated from cucumber. MGDG may be	
XX	CC	used to increase lipid production in an organism.	
XX	SS	Sequence 422 AA;	
Qy	5	KRVLIILTANYGNHG---VQVAKTILYEQCRVLGFQHVTVSNLYQESNPVISEVTQYLKLS	61
Query Match		15.5%; Score 305.5; DB 19; Length 422;	
Best Local Similarity		25.7%; Pred. No. 7.9e-21;	
Matches	100;	Conservative 78; Mismatches 178; Indels 33; Gaps 13;	

Db	33	krvllmsdtggghrasaeaikaafneefgnnyq-vftldwtcdtpwfnqlprsynfl	91
Qy	62	FSIGKQYRLFYGYGVDDIYNKRKFNIYFKMGKRLGELVDEHOPDIINFPW----	116
Db	92	vkhgltwkmtyyvtapkvihgsnfaaststfiarevakglmkypdliisvhlmgvhpir	151
Qy	117	VPEYRRTRGRVIPFNVTDF-CLHKITWHEHNDKYIVADYYKEKLLEIGTHPSNVKIT	175
Db	152	ilrskgllnkivft-tvtdlstchptwfklvltrcycpstevakalkaglpqslkfv	210
Qy	176	GIPRPOFEESMPVGP---TVKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKD	229
Db	211	glvrpsfvk--piipkielrkclgmndenlpavlingggemgpieatkaalskalyden	268
Qy	230	Q-----VOVVVGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMTTKPGGIT	285
Db	269	hgpeigqvlvicghnkklagrlrsidwk--vpvgkgfvtkmecmgacdcitkagpgt	326
Qy	286	LTEATATGVPVILYKYPGQEKENAFEDRGAAIVVNRHEEILESVTSLLADE-DTLHR	344
Db	327	iaeamirgpliiindyiaaggaenvpyvvengcgkfkskpkeianivakwfgpkadelll	386
Qy	345	MKNIKDLHLANSSEVI-----LEDILKE	368
Db	387	msqna--lrarpdavfkivhdlhelvkq	413
RESULT	6		
AAB19040			
ID	AAB19040	standard; protein; 525 AA.	
AC	AAB19040;		
XX			
DT	08-FEB-2001	(first entry)	
XX			
DE		Amino acid sequence of monogalactosyl-diacetylglucosyl (MGDG) synthase.	
KW		Monogalactosyl-diacetylglucosyl synthase; MGDG synthase; Toxoplasma;	
KW		apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;	
KW		acquired immune deficiency syndrome; coccidiosis.	
OS		Cucumis sativus.	
XX			
PN	WO200056919-A1.		
PD		28-SEP-2000.	
PF	17-MAR-2000; 2000WO-FR00658.		
PR	19-MAR-1999; 99FR-0003434.		
PA	{COMS } COMMISSARIAT ENERGIE ATOMIQUE.		
PI	Marechal E, Block M, Joyard J, Douce R;		
WI	WPI; 2000-602227/57.		
PT	Use of monogalactosyl diacetylglucosyl synthase for identifying its		
PT	specific inhibitors, potentially useful as antiparasitic agents and		
XX	herbicides -		
XX			
XX	Example 2; Fig 2; 33pp; French.		
CC	The present sequence represents a monogalactosyl-diacetylglucosyl (MGDG)		
CC	synthase. MGDG is present in all plastids tested and is essential for		
CC	cell survival, but is not present in other membrane structures, or in		
CC	animal cells, and so represents a specific target. MGDG synthase or a		
CC	plastid membrane isolated from plants is used to select and screen for		
CC	specific inhibitors of MGDG synthase. These inhibitors are suitable as		
CC	active agents against apicomplex parasites or as herbicides. The		
CC	inhibitors are used as antiparasitic agents, especially against		
CC	Plasmodium, Toxoplasma (particularly in patients with acquired immune		
CC	deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry).		

CC as herbicides.
XX
SQ Sequence 525 AA;

Query Match 15.4%; Score 303.5; DB 21; Length 525;
Best Local Similarity 25.7%; Pred. No. 1.7e-20;
Matches 100; Conservative 77; Mismatches 179; Indels 33; Gaps 13;

QY 5 KRVLLTANYGNH--VQVAKTYECCVRLGFOHVTVSNLYOESNPVSEVTQYLKLS 61
DB 136 krvlmsdtdggghrasaakaafneefgnnyq-vftdldtthtpwfnqprsynfl 194
QY 62 FSGIKQRYLRYFYGVVDKIYKRNKRIYFKMGNKRLGELVDEHQPDIINTFPM 116
DB 195 vkhtlwmkyvvtapkvihqsnfaaststfiarevakglmkypdliisvhpImqhvpir 254
QY 117 VPEYRRRTGVRIPFNVTDF-CLHKTIWVHENVDKYVATDYVKEKLEIGTHPSNVKIT 175
DB 255 ilrskgllnkivft-tvtdlstchptwfhkvtircycpstevakraglqpsklkvf 313
QY 176 GIPIRPFESMPVGP---IYKYNLSPNKKVLLIMAGAHG---VLKNVKELCNLVKDD 229
DB 314 glvrpsfvk--prrpkvelrrelgmdenlpavllmggggmgpieatakalskalyden 371
QY 230 Q-----QVVVCGKNTALKESLSALEANGDKLVGLYVERIDELEFRITDCMITKPGGIT 285
DB 372 hgepqlvqlvicghnklagrlrsldwk--vpvqkvgtkmeecmgacdciitkagpgt 429
QY 286 LTEATAGVPIVLYKVPVGOEKENANFFEDRGAIVVNRHHEILESVTSLADP-DPLHR 344
DB 430 iaeamirglpilndylagqagvpyvngcgkfskpskeianivakwfgpkadel11 489
QY 345 MKNKIKDLHLANSSEVI-----LEDILKE 368
DB 490 msqna--lrlarpdavfkivhdhclvkvq 516

RESULT 7
AAB19041
ID AAB19041 standard; protein; 533 AA.
AC AAB19041;
DT 08-FEB-2001 (first entry)
XX Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
DE Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW apicomplex parasite; herbicide; antiparasitic; Plasmodium; Elmeria;
XX acquired immune deficiency syndrome; coccidiosis.
OS Arabidopsis thaliana.
PN WO200056919-A1.
XX 28-SEP-2000.
XX 17-MAR-2000; 2000WO-FR00658.
XX 19-MAR-1999; 99FR-0003434.
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX Marechal E, Block M, Joyard J, Douce R;
PI WPI; 2000-60227/57.
XX use of monogalactosyl diacylglycerol synthase for identifying its
XX specific inhibitors, potentially useful as antiparasitic agents and
XX herbicides
XX Example 2; Fig 2; 33pp; French.

XX The present sequence represents a monogalactosyl-diacylglycerol (MGDG) synthase. MGDG is present in all plastids tested and is essential for cell survival, but is not present in other membrane structures, or in animal cells, and so represents a specific target. MGDG synthase or a plastid membrane isolated from plants is used to select and screen for specific inhibitors of MGDG synthase. These inhibitors are suitable as active agents against apicomplex parasites or as herbicides. The inhibitors are used as antiparasitic agents, especially against Plasmodium, Toxoplasma (particularly in patients with acquired immune deficiency syndrome) or Elmeria (coccidiosis in cattle or poultry), and as herbicides.

XX
SQ Sequence 533 AA;

Query Match 15.3%; Score 301.5; DB 21; Length 533;
Best Local Similarity 24.8%; Pred. No. 2.7e-20;
Matches 101; Conservative 86; Mismatches 172; Indels 49; Gaps 17;

QY 5 KRVLLTANYGNHGVQVAKTYEQCVRLGFO-----HVTVSNLYOESN--PIVSEVTQ 55
DB 142 kkvllmsdtdggghrasa-----eairaaafngedgdyqvftdltwcdhtpwnfnqlprs 196
QY 56 YLKLKSFSGIKQRYLRYFYGVVD-KIYNKRNKRIYFKMGNKRLGELVDEHQPDIINTFPM 114
DB 197 ynfvlkvgh--tlwmkyvgtsprihqsnfaaststfiarevakglmkypdliisvhp1 253
QY 115 I--VVPYRRRTG--RVIPFNVTDF-CLHKTIWVHENVDKYVATDYVKEKLEIGTHP 169
DB 254 mqhvplrvltskglkllkvftvtdlstchptwfhkvtircycpstevakraglqaglet 313
QY 170 SNKVTGIPRPFESMPVGP---IYKYNLSPNKKVLLIMAGAHG--LKNVKELCN 224
DB 314 sqkvvyglvrpsfvk--prrpkvelrrelgmdenlpavllmggggmgpieataralad 371
QY 225 LVKDDQV-----QVVVCGKNTALKESLSALEANGDKLVGLYVERIDELEFRITDCMIT 279
DB 372 alvdknlgeavgvliicgrnklqsklssldwk--lpvqkvgtkmeecmgacdciit 429
QY 280 KPGGITLATEATAGVPIVLYKVPVGOEKENANFFEDRGAIVVNRHHEILESVTSLAD- 338
DB 430 kagpgtiaeamirglpilndylagqagvpyvngcgkfskpskeianivakw 485
QY 339 -----EFTLHRMKNKIKDLHLANSSEVILEDILKESEMMTAKQAKVLS 382
DB 486 fgpaskeleimsqna--lrla-kpeavfkivhdhclvkrkkslpqls 530

RESULT 8
AAB19042
ID AAB19042 standard; protein; 468 AA.
AC AAB19042;
DT 08-FEB-2001 (first entry)
XX Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
DE Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW apicomplex parasite; herbicide; antiparasitic; Plasmodium; Elmeria;
XX acquired immune deficiency syndrome; coccidiosis.
OS Arabidopsis thaliana.
PN WO200056919-A1.
XX 28-SEP-2000.
XX 17-MAR-2000; 2000WO-FR00658.
XX 19-MAR-1999; 99FR-0003434.
XX 99FR-0003434.

(COMS) COMMISSARIAT ENERGIE ATOMIQUE.

Marechal E, Block M, Joyard J, Douce R;

WPI; 2000-602227/57.

Use of monogalactosyl diacylglycerol synthase for identifying its specific inhibitors, potentially useful as antiparasitic agents and herbicides

Example 2; Fig 2; 33pp; French.

The present sequence represents a monogalactosyl-diacylglycerol (MGDG) synthase. MGDG is present in all plastids tested and is essential for cell survival, but is not present in other membrane structures, or in animal cells, and so represents a specific target. MGDG synthase or a plastid membrane isolated from plants is used to select and screen for specific inhibitors of MGDG synthase. These inhibitors are suitable as active agents against apicomplex parasites or as herbicides. The inhibitors are used as antiparasitic agents, especially against Plasmodium, Toxoplasma (particularly in patients with acquired immune deficiency syndrome) or Elmeria (coccidiosis in cattle or poultry), and as herbicides.

Sequence 468 AA;

Query Match 14.9%; Score 293; DB 21; Length 468;

Best Local Similarity 25.1%; Pred. No. 1.5e-19;

Matches 98; Conservative 86; Mismatches 172; Indels 34; Gaps 14;

QY 5 KVLVLTANYGNHGVQVAKTYEQC-VRLGFOH-VTVSNLYOESN--PIVSEVTQYLYLK 60

Db 69 knvllmsdtggghrasaeairdafkiefgkyrvivkdvwkeytgwplndmersyktmv 128

QY 61 SPSIGKQYRLPYG-----VDKIYKRNKFNIFKMGKRLGELVDEHQPDIINTFPM- 114

Db 129 kh---vqlkvafstspkwhscylaaiaayakeveagl----meykpelilsvhplm 181

QY 115 ----IWPPEYRRRTGRVPTFNMVDF-CLHKIWHENVNRYVATDYVKEKLEIGTHP 169

Db 182 qhpiwlkwqelqkrvi-fvtvitdltchptwfhgpnrcycpsgevakralfdgide 240

QY 170 SNVKTGPIRPFQESMPV-GPIYKKNLSPNKKVLLIMAGHGV--LKNVKELCENLV 226

Db 241 sqrvrfigpvrpsfarvlvkdldirkelemdqdlravllmggggmgpvketakalef1 300

QY 227 KDDQ-----VOVVCGKNTAKESLSALEANGDKLVGVVERIDELFRITDCMITKP 281

Db 301 ydkenrkpiqmvvicgrnkklaaleaidwk--lpvkvrgetqmekwmgaacdciitka 358

QY 282 GGITLATEATGVPVILYKVPVGGQKENANFFEDRGAAIVVNRHHEILESVTSLIADP-D 340

Db 359 gpgtiaelirslpilndyipgkgnvpyvengagvfrspketarivgfwfstkttd 418

QY 341 TLHRMKNKIDLHLANSSEVILEDLKESE 370

Db 419 electsdnarklaqpeavfdvkdideise 448

RESULT 9

AAG42415

ID AAG42415 standard; Protein; 404 AA.

AC AAG42415;

XX 18-OCT-2000 (first entry)

DT Arabidopsis thaliana protein fragment SEQ ID NO: 52895.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52895.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hydridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.6%; Score 287.5; DB 21; Length 404;
Best Local Similarity 23.6%; Pred. No. 3.9e-19;
Matches 97; Conservative 88; Mismatches 169; Indels 57; Gaps 17;

Qy 15 NGHVQVAKTL-----YEQCVRLGFO---HVTVSNLYQESN--PIVSE 52
Db 5 99ghrasaeairaafnqefgdeyqsglefelelekgfrclvhvftldtwdhtpwpfnql 64
Qy 53 VTQYLYLKSFSIGKOFYRLFYGVGD-KIYNKRKNFIYFKMGKNKLGELVDEHQDPIINT 111
Db 65 prsynflvkhg---tlwkmtygtsprivhdsnfatstfiareiaqglmkypqdilsv 121
Qy 112 FPMI--VYPEYRRRTG--RVIPTFNMTDF-CLHKIWIHENVNVDKYVATDYVYKELLETG 166
Db 122 hplmqhvpvlrvlrskgllkkivftvtidlstcptwfhkvltrcycpstevakraqag 181
Qy 167 THPSNVKITGPIRPOFESMPVGP---IYKKNLSPKKVLLIMAGAHGV--LKNVKEL 221
Db 182 letsqikvygipvrpsfvk--pvrpkvelrrelgmdenlpavllmgggmgpieatara 239
Qy 222 CENLVKDDQV-----QVYVCGKNFALRESLSALEANGDKLVLGYVERIDELFRITDC 276
Db 240 ladalydknlgeavqgvliicgrnkklsksslawk--ipvqvkfittkmeecmgadc 297
Qy 277 MITKPGGITLFEATAIGVPIVILYKVPQGEKENANFFEDRGAAIYVNRHEEILESVTSL 336
Db 298 litkagpgtiaeamirgipillingyiaqgeagnvpyvvengc----gfkspkskiskiv 353
Qy 337 AD-----EDTLHRMKNKIKDLHLANSSEVILEDILKESEMMTAKOKAKVLS 382

Matches 80; Conservative 71; Mismatches 158; Indels 85; Gaps 18;

```
QY 6 RVLILITANGNGHVQVAKTLVQECVRLCFQHVTVSNLQESNPIVSE--VTQYLKLSFS 63
Db 4 kcfmimgagtgnifpalavads-lrvrghnv-----lwgsksmeerlvpqy----- 51
QY 64 IGKQFYRLFYGVGDKIYNNKRFNIYFKMGN--KRLGELVDEHQPDIIIN-----TFP--- 113
Db 52 -girlletlaigirngikrklmlpftlyktvreaqriihrkhrvecvigfgvftfpogl 110
QY 114 -----MIVPEYRRRG---RVPTFNMTDFCLHKIWHVNDKYVYAFDYVKEKLL 163
Db 111 aaklgvpiivtheqavaglsnrhlrsrwakrvlyafpkafshg----- 154
QY 164 EIGHTSPNVKTINGIPRQFEESMPVGLYKYNLSPNKKVLLI--MAGAHGLVKNVKEL 221
Db 155 -----givnnpvradi-snipv-paerfqggregrlkilvvgsglgadvlnktvpqa 203
QY 222 CENLVKDDQVQVVVCGKNKTALKESLSALEAEN---GDKLVGLGYVERIDELFRITDCMI 278
Db 204 lallpeevrpgmyhqsgrn-----klgnldadydalgvkaecvefitdmvsayrdadivi 258
QY 279 TKPGGITLATEAIGVPVILKVPV-----GOEKENANFFEDRGAAIYVNRHEEILSVTS 334
Db 259 cragaltiaeltaagiallv-pyphavddhgtanarfmvqaeaglllpqqltaeklae 317
QY 335 LIADEDTLHRMK-----KNIKDLHLANSSEVILE 363
Db 318 ilg---sinrekclwaenartlalphaseadvaee 348
```

RESULT 12

AAW98240
ID AAW98240 standard; Protein; 993 AA.

AC AAW98240;

DT 31-MAR-1999 (first entry)

DE H. pylori GHPO 741 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.

OS Helicobacter pylori.

PN WO9843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

PA (HUMA-) HUMAN GENOME SCI INC.

PI (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

DR WPI: 1998-542293/46.

DR N-FSDB; AAX13959.

XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases

XX Claim 8; Page 200-204; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with

CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.

XX Sequence 993 AA;

Query Match 5.7%; Score 112.5; DB 19; Length 993;
Best Local Similarity 20.5%; Pred. No. 0.091;

Matches 95; Conservative 70; Mismatches 154; Indels 145; Gaps 21;

```
QY 6 RVLILITANGNGHVQVAKTLVQECVRL-----GFOHVTVSNLQESNPIV 50
Db 101 lliiwektdkgeknngikdikeqsifireiplmtertsfiingvervvvqnhrspgvlf 160
QY 51 SEVTQYLYLKFSFGKQFYRLFYGVGDKIYNNKRFNIYFKMG-----NKR----- 95
Db 161 keee-----sstsink-----lytg--qilpdrsgwlyfeydskdvlyarinkrrkvpvt 209
QY 96 -LGEIVDEHQPDIIINTFPMIVVPEYRRRTGRVPTFNMTDFCLHKIWHVNDKY--- 151
Db 210 ilframdyqgqdlkkmfplvkvr-----yen-dkylip 242
QY 152 YVATDYVKEKLEIGCTHPSNV-----KITGIPRQFEESM-----PVGPIYKYNLSP 200
Db 243 fasidanqrmeffdikdpqgkvillagkkltsrkkelkenhleweyepmdillnrhlaep 302
QY 201 NKKVLLIMAGAHGLVKNVKELCEN---LVKDDQVQVVV---CGKNTALKESLSALEA 252
Db 303 -----vmvgkevildmitqldknklekihdilgvgefviindlalghdasiqfsa--- 353
QY 253 ENGDKLVGLGYVERIDE---LFRITDCMITKPGGITLATEAIGVPVILYKP-----VP 303
Db 354 -dseslklkqtteidenaalairihkvmkpgdpvttevakqfvkkliffoperydlcmv 412
QY 304 GOEKENANFFEDRGAAIYVNRHEEI-----LESVTSLL 336
Db 413 grkmnhk1glhvpdyittlthediittvkvylmkiknnqgkiddrdhlnriraavgell 472
QY 337 ADEDTLH-----RMKKNIKD--LHLANSSEVILEDLKESEMMTA 374
Db 473 ane--lshglvmkqtkldkittmsgafdsimphdlvnskmlts 514
```

RESULT 13

AAAR20198
ID AAR20198 standard; Protein; 1068 AA.

AC AAR20198;

DT 14-APR-1992 (first entry)

DE Sucrose phosphate synthase from corn.

XX SPS; fructose-6-phosphate; UDP-glucose; sucrose regulation;
KW carbon partitioning; plants.

OS Zea mays L. cv Pioneer 3184.

PH Key Location/Qualifiers
FT Peptide 71..74
FT /label= A8

FT /note= tryptic "

FT Peptide 206..212

FT /label= B4

FT /note= tryptic "

FT Peptide 471..481

FT /label= B11

FT /note= tryptic "

FT Peptide 872..884

FT /label= 4K

FT /note= tryptic "

```

FT Peptide 885..891
FT /label= 12N
FT /note=" tryptic "
PN EP466995-A.
XX 22-JAN-1992.
XX 20-JUL-1990; 90EP-0402084.
XX 20-JUL-1990; 90EP-0402084.
XX (ROUS ) ROUSSEL UCLAF.
XX Van Assche C, Lando D, Bruneau JM, Voelker TA, Gervais M;
XX WPI; 1992-025895/04..
XX N-PSDB; AAQ20664.
XX Sucrose phosphate synthase (SPS) from corn - its prepn., DNA
XX encoding it, and hybridomas and monoclonal antibodies specific for
XX it.
XX Claim 5; Fig 7; 40pp; English.
XX The sequence, which is that of sucrose phosphate synthase (SPS),
XX was deduced from a cDNA sequence obtd. from three clones isolated
XX from a cDNA library prepd. from RNA isolated from corn leaves. The
XX tryptic peptides (see features) were used to design degenerative
XX oligonucleotide primers and probes (see AAQ21987-93) for the prepn.
XX by PCR of a probe for the SPS gene. The gene can be used to prepare
XX a vector for expression of recombinant SPS, which is a key enzyme
XX in sucrose regulation and carbon partitioning between starch and
XX sucrose in photosynthesising plants. (Plants transformed with the
XX DNA have an increased flow of sucrose to growing tissues and give
XX increased yields). The recombinant SPS can be used to raise mono-
XX clonal antibodies which can then be used to purify the protein by
XX affinity chromatography.
XX Sequence 1068 AA;

Query Match 5.7%; Score 112.5; DB 13; Length 1068;
Best Local Similarity 20.9%; Pred No. 0.1;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;

QY 125 GRVPTFNVM---TDFCLHKIWHENVKYYVATDYVKEKLEI-GTHPSNVKITGPIR 180
DB 427 grypmrmvippgmdf--snvvvhedid----gdgkvkddivlegaspkmp----- 473
QY 181 PQFEESM-----PVGPIYKYNLSPNKKVLLIMAGAGHVLKNVKELCENLVKDDQVQVV 235
DB 474 plwaevmrfllphkpmilalsrpdpkntttlvkafgecrpirl-----anltl 524
QY 236 VCGKNTALKESLSALEANGDKL-KVLGYVERID-----ELFRITDCM 277
DB 525 imgn-----rddidmsagnasvltttvklidkydlygsvafphkhnqadvpeirylaakm 580
QY 278 -----ITKPGGILTEATAIGVPVILYKVPQGEKENANFFEDRGAAYVNRHEE-- 327
DB 581 kgvfnpalvepfgltlieaahgipivatknngpvditna-----lmgllivdpdqna 635
QY 328 ILESVTSLLADEDTLHRMKN-KIKDLHL 354
DB 636 iadallklvadknlwqecrringlrnlh 663

RESULT 14
AAW38266
ID AAW38266 standard; Protein; 1068 AA.
XX
AC AAW38266;
XX

```

```

DT 11-MAY-1998 (first entry)
XX
XX Maize sucrose phosphate synthase.
XX
XX Sucrose phosphate synthase; SPS; transgenic plant; maize; corn.
XX
XX Zea mays L. cv. Pioneer 3184.
XX
XX Key Location/Qualifiers
XX Peptide 71..74
XX Peptide /label= A8
XX Peptide 206..212
XX Peptide /label= B4
XX Peptide 471..481
XX Peptide /label= B11
XX Peptide 872..885
XX Peptide /label= 4K
XX Peptide 886..892
XX Peptide /label= 12N
XX
XX EP807685-A2.
XX
XX 19-NOV-1997.
XX
XX 20-JUL-1990; 90EP-0402084.
XX
XX 20-JUL-1990; 90EP-0402084.
XX
XX 20-JUL-1990; 90EP-0201062.
XX
XX (ROUS ) ROUSSEL-UCLAF.
XX
XX Bruneau J, Gervais M, Lando D, Van Assche C, Voelker TA;
XX WPI; 1997-552360/51.
XX N-PSDB; AAT95847.
XX
XX DNA encoding sucrose phosphate synthase - useful for producing
XX transgenic plants
XX
XX Claim 5; Fig 7; 38pp; English.
XX
XX This protein comprises maize sucrose phosphate synthase (SPS), an
XX enzyme that catalyses the formation of sucrose phosphate from
XX fructose 6-phosphate and UDP-glucose in photosynthetically active
XX plant cells. It is considered to be a rate-limiting enzyme in the
XX pathway providing sucrose to growing tissue. The SPS amino acid
XX sequence was deduced from cDNA clones (see AAT95847) derived from
XX leaf cDNA. The complete protein is believed to be a dimeric or
XX tetrameric protein having a basic subunit of 110-130 kDa. Peptides
XX B11 (see AAW38269) and 4K (see AAW38270) isolated from SPS were used
XX to design primers (see AAT95848-49) utilised in the isolation of
XX SPS cDNA. Transgenic plants that express SPS can be used to
XX measure the effects on crop yield of an increased rate of sucrose
XX translocation to growing tissues.
XX
XX Sequence 1068 AA;

```

```

Query Match 5.7%; Score 112.5; DB 18; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;

QY 125 GRVPTFNVM---TDFCLHKIWHENVKYYVATDYVKEKLEI-GTHPSNVKITGPIR 180
DB 427 grypmrmvippgmdf--snvvvhedid----gdgkvkddivlegaspkmp----- 473
QY 181 PQFEESM-----PVGPIYKYNLSPNKKVLLIMAGAGHVLKNVKELCENLVKDDQVQVV 235
DB 474 plwaevmrfllphkpmilalsrpdpkntttlvkafgecrpirl-----anltl 524
QY 236 VCGKNTALKESLSALEANGDKL-KVLGYVERID-----ELFRITDCM 277
DB 525 imgn-----rddidmsagnasvltttvklidkydlygsvafphkhnqadvpeirylaakm 580

```

```
QY 278 -----ITKPGGITLTAATGVPVILYKVPVGOEKENANFFEDRGAAlVVRHEE-- 327
Db 581 kgvfinpalvepfgltlieaaahglpivatknngppvdtna-----lnngllvdphdqa 635
QY 328 ILESVTSLLADEDTLHRMKN-IKDLHL 354
Db 636 iadallklvadknlgwecrrnglrnhl 663

RESULT 15
AAW09869
ID AAW09869 standard; Protein; 1068 AA.
AC AAW09869;
XX
DT 25-JUL-1997 (first entry)
DE
XX
XX Sucrose phosphate synthase.
KW Sucrose phosphate synthase; sweetness; transgenic plant; tomato;
KW fruit; tuber; sugar; maize; corn.
XX
OS Zea mays hybrid 3184.
XX
PN WO9715678-A2.
PD
XX
XX 01-MAY-1997.
XX
XX 25-OCT-1996; 96WO-US17351.
XX
XX 27-OCT-1995; 95US-0549016.
XX
XX (CALJ ) CALGENE INC.
XX
XX Shewmaker CK;
XX
XX WPI; 1997-259030/23.
DR N-PSDB; AAT66229.
XX
XX Modifying sweetness of plant parts by introducing transgene that
PT encodes sucrose phosphate synthase -- also altering ratio of soluble
PT solids in sink tissue, especially for increasing sugar content in
PT tomatoes
XX
XX Example 5.6; Page 69-75; 112pp; English.
XX
XX The amino acid sequence (AAW09869) of maize hybrid 3184 sucrose
CC phosphate synthase (SPS) was deduced from cDNA clones (see also
CC AAT66229) isolated from a leaf cDNA library. SPS is considered a
CC rate-limiting enzyme in the pathway providing sucrose to growing
CC tissue. Methods are provided for modifying the sweetness of plant
CC sink tissue (partic. tomato fruit) in which SPS activity and/or
CC invertase activity in plant tissues is manipulated. Carbohydrate
CC partitioning can be modified in plant tissues and/or parts, which
CC in turn can be used to alter plant growth, soluble solid content
CC and/or sweetness, and/or to alter the sensitivity of plant growth
CC to temperature and/or to levels of carbon dioxide and oxygen.
XX
XX Sequence 1068 AA;
```

```
Query Match 5.7%; Score 112.5; DB 18; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;

QY 125 GRVTPTEENV---TDFCLHKIWHENVKYVATDVYVKELLET-GPHPSNVKITGIPR 180
Db 427 grymtrmvippgmdf--snvvvnedid----gdgdvkddivglegaspkmp----- 473
QY 181 PQFEESM-----PVGPIYKKNLSPNKKVLLIMAGAHGVLKNVKELCNLVKDDQVQVV 235
Db 474 piwaevmrftnphkpmilalsrpdpkknittivkafgecrpirel-----antli 524
```

```
QY 236 VCCKNTALKESLSALEAENGDKL-KVLGYVERID-----ELFRITDCM 277
Db 525 imgn-----rddiddmsagnasvlttvlklldkydlygsvafpkhngadvpeiylaakm 580
QY 278 -----ITKPGGITLTAATGVPVILYKVPVGOEKENANFFEDRGAAlVVRHEE-- 327
Db 581 kgvfinpalvepfgltlieaaahglpivatknngppvdtna-----lnngllvdphdqa 635
QY 328 ILESVTSLLADEDTLHRMKN-IKDLHL 354
Db 636 iadallklvadknlgwecrrnglrnhl 663
```

Search completed: June 29, 2001, 08:59:02
Job time: 175 sec

3

QY 178 PIRQFESMPVGPYIKKY-----NLSPNKKVLLIMAGAHGVKLVKNVKELCEN-LVKDDQV 231
Db 178 PIDNKEET-----PINQKWLIDNLDPKQTILMSAGAFGVSGFDTMITDILAKSANA 232
QY 232 QVVVVCNKTALKESLSALEANGDKLK-----VLGYVERIDELFRITDCMTKPGGIT 285
Db 233 QVVMICGSKELKSLTA-----KFLTRMVLILGYTKHMNEWMASSQLMITKPGGIT 285
QY 286 LTEATAJGVPIVLYKVPQGEKENANFFEDRGAIVVNRHEEILSVTSLLADEDTLHRM 345
Db 286 ITEGFARCIPIWFLNPPAGGELENAFEEKGGKGIADTPEBAIKIVASLTNGNEQLTNM 345
QY 346 KKNIKDLHLANSSEVILEDL 366
Db 346 ISTMEQDKIKYATQICRDLL 366

RESULT 2
Q9KBH0 PRELIMINARY; PRT; 374 AA.
AC Q9KBH0
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (T-EMBLrel. 15, Last annotation update)
DE BH1957 PROTEIN.
GN BH1957.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=8665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001513; BAB05676.1; --
SQ SEQUENCE 374 AA; 42361 MW; DA96A0EA33C5AC65 CRC64;

Query Match 18.7%; Score 368; DB 2; Length 374;
Best Local Similarity 27.0%; Pred. No. 6.1e-16;
Matches 103; Conservative 84; Mismatches 159; Indels 36; Gaps 13;
QY 8 LILITANYGNHGVQVAKLYEQCVRLGFQHVTVVSNLQESNPVISEVTOYLKSFSGIKQ 67
Db 6 LIFSASIGNGHQAALQALQVEFQNGYQPEIIDTFYSLS-PALHKFMLTYSVNLKVGPR 64
QY 68 FYRLFYGVDK---IYNKRKNFYKMGKRLGELVDEHOPDIIINTFPMIVVPEYRRRT 124
Db 65 IWQKIYFOAEKYPFLFLDQFATFFV---ESLHATVKSNCRSFLVSTHPVTAFLVRLKS 121
QY 125 GRV--IPTFNMTDFCLHKIWHVHENVKYYVA---TDVYKELLETGTHPSNVK----- 173
Db 122 KQQLNPLVYITDFVLHPAYLRPEIDGYTSDPNFTDAK-----LNNVSDREF 172
QY 174 ITGIPTRQPEE-SMPVGPIYKYNLSPNKKVLLIMAGAHGVKLVKNVKELCENI-VKDDQV 231
Db 173 PTGIPT-PNLESIDQPKWKVRNDGLDQPKVLIAGGGIG-LTNYAQVIRALECLPEPI 230
QY 232 QVVVVCNKTALKESLSALEANGDKLVGYVERIDELFRITDCMTKPGGITLFEATA 291
Db 231 QLCMIGHNYQVREKISRKSKB-ELKVTEFTDFKLLYKASDAILSRAGGLTMAESLV 288
QY 292 IGVPIVLYKVPQGEKENANFFEDRGAIVVNRHEEILSVTSLLADEDTLHRMKNKID 351
Db 289 CETPIIHQVPVPGHEHNKFLIDAGAAALRVKSGKEIPTTKRVLYEACFGPMIENARK 348
QY 352 LHLANSSEVILEDL---KESE 370
Db 349 LKPNAAANEIVEQMLLVKEQQ 370

RESULT 3

Q9RVF3 PRELIMINARY; PRT; 411 AA.
AC Q9RVF3
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE CELL WALL SYNTHESIS PROTEIN, PUTATIVE.
GN DR1076.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamatheva J.J., Lam P., McDonald L., Otterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL: AE001958; AAF10649.1; --
DR TIGR: DR1076;
SQ SEQUENCE 411 AA; 45017 MW; D825DE52B8801437 CRC64;

Query Match 17.9%; Score 353; DB 2; Length 411;
Best Local Similarity 26.2%; Pred. No. 6.3e-15;
Matches 101; Conservative 77; Mismatches 164; Indels 44; Gaps 11;
QY 6 RVLLITANYGNHGVQVAKLYEQCVRLGFQHVTVVSNLQESNPVISEVTOYLKSFSG 63
Db 29 RALFMSVSLGAGHDQA-----QOAVKQAFARGVGLLGAHDSV-----EYLSTFERSFT 78
QY 64 I-----GKQFYRLFYGVDKIYNKRKNFYKFM---GNKRLGELVDEHOPDIIIN 110
Db 79 VDLVYFELRYAPWLYRGFWLTQD---DQPNLIISRMFTWLGMAFKDELRLPEVIN 135
QY 111 TP--PMIVPEYRRRTGRVPTFNMTDFCLHKIWHVHENVKYYVATDYVYKELLETGTH 168
Db 136 SFWAPAACVDTLRAQTQGRFLNCLIVTDYRAHLHARRETDLMLVASEETRQMLRGVR 195
QY 169 PSNVKITGPIRQPEESMPV-----GPIYKYNLSPNKKVLLIMAGAHGVKLVKNKELC 222
Db 196 PEQVEVIGIPISPAFREVLAADRWALRAELFSEMSLRPGVPLLLLSGGGRGHYAAADV 255
QY 223 ENLVK-DDQVQVVVVCNKTALKESLSALEANGDKLVGYVERIDELFRITDCMITKP 281
Db 256 TELGNLGRAVQVLPASRQEGTETI-----GGAIVHHLGFRDLPRLLAASDLVVGKA 309
QY 282 GGITLFEATAJGVPIVLYKVPQGEKENANFFEDRGAIVVNRHEEILSVTSLLADEDT 341
Db 310 GGLTVAEATAGVPLVIYAPIQGEENADFLERHAGLWARAHDRVRLVLRAL-DPAE 368
QY 342 LHRMKNKIDLHLANSSEVILEDLK 367
Db 369 HARLSAGARAVGIPDAADRVRGAILR 394

RESULT 4
Q9SM44 PRELIMINARY; PRT; 522 AA.
AC Q9SM44
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE MGDG SYNTHASE A PRECURSOR (EC 2.4.1.46).
GN MGD A.
OS Spinacia oleracea (Spinach).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
OC Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99449603; PubMed=10518794;
RA Mele C., Marechal E., Shimojima M., Awai K., Block M.A., Ohta H.,
RA Takamiya K.I., Douce R., Joyard J.;
RT "Biochemical and topological properties of type A MGDG synthase, a
RT spinach chloroplast envelope enzyme catalyzing the synthesis of both
RT prokaryotic and eukaryotic MGDG.";
RL Eur. J. Biochem. 265:990-1001(1999).
DR EMBL: AJ249607; CAB56218.1; -
DR TRANSIT 1 98
FT CHAIN 99 522 MGDG SYNTHASE A.
SQ SEQUENCE 522 AA; 57511 MW; 02E2B929732551A7 CRC64;

Query Match 16.5%; Score 326; DB 10; Length 522;
Best Local Similarity 25.3%; Pred. No. 4.6e-13;
Matches 98; Conservative 80; Mismatches 162; Indels 48; Gaps 12;

QY 5 KRVLITANYGNHGVQVATLYEQVRLGQ-----HVTVSNLQESN--PIVSEVTO 55
Db 131 KRVLITANYGNHGVQVATLYEQVRLGQ-----HVTVSNLQESN--PIVSEVTO 55
QY 56 YLYLKSFSIGKQYRIFYGYVD-KIYNKRKFNIFKMGKRLGELVDEHOPDIINTFPM 114
Db 186 YNPLVKHG---PLWKMYGTSPRVHQSNFATSVFIAREVARGMLKQPDIIISVHPL 242
QY 115 IVVPEYRRRTGR---VIPTFNMTDF-CLHKIWHVHNDKYVATDYVYKELLEGTHP 169
Db 243 MQHVLRLRGLLEKIVFTTVDLTSTCHPTWPHKLVTRCYCPSNEVAKRATKAGLP 302
QY 170 SNVKTGPIRQFESM-PVGPYIKYNLSPNKKVLLIMAGHGV--LKNVKELCENLV 226
Db 303 SQIKVGLPVSFVSRLKELRKEKMDHEPLPAVLLMGEGGEGMGPTEARALGNAL 362
QY 227 KDOV-----QVVVCGKNTALKESLSALEANGDKLVGYVERIDELFRITDCMITKP 281
Db 363 YDANLEPTGQLLVICGRNKKLAGLSSIDWK--IPVQVGFYTKIECGMGACDCIITKA 420
QY 282 GGTTFEATAIGVPVILYKVPQGEKENANFFEDRGAIVVNRHIELESVTSLLADEPT 341
Db 421 GPGTIAEAMIRGLPIILNDYIAGQAGNVPYVIENGIGYKSPKEIAKTVSQWFGPK-- 478
QY 342 LHRMKNKIDHLANSSEVILEDKES 369
Db 479 -----ANELQMSQNALKHA 493

RESULT 5
Q9SI93 PRELIMINARY; PRT; 464 AA.
AC Q9SI93;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE PUTATIVE MONOGLACTOSYLACYLGLYCEROL SYNTHASE.
GN AT2G11810.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC007187; AAD28678.1; -
DR InterPro: IPR001296; -
DR Pfam: PF00534; Glycos_transf_1; 1.
SQ SEQUENCE 464 AA; 52859 MW; FEC2B424CFBBA136 CRC64;

Query Match 16.1%; Score 317; DB 10; Length 464;
Best Local Similarity 25.6%; Pred. No. 1.5e-12;
Matches 101; Conservative 90; Mismatches 171; Indels 32; Gaps 16;

QY 5 KRVLITANYGNHGVQVATLYEQVRLGQ---QHVTVSNLQESN--PIVSEVTOYLYL 59
Db 72 KTVLILMSDTGGHRAAEAI-RDAFKIEFGDDYRIIRKDVWKEYTGWPLNDMERQYKFM 130
QY 60 KSFSGIKQFYRIFYGYVD-KIYNKRKFNIFKMGKRLGELVDEHOPDIINTFPM--- 114
Db 131 VK-HVG--LWSVAFHCTSPKWIHKSYLSALAAAYAKEIEAGLMEYAPDIIISVHPLMQHI 187
QY 115 -IVVPEYRRRTGRVPTFNMTDF-CLHKIWHVHNDKYVATDYVYKELLEGTHPSNV 172
Db 188 PLWVMKWOGLHKKVI-FVTVITDLNCTHRTWPHHGVSRVCYCPSEKAKRALVDGLDSOI 246
QY 173 KITGPIRQFESM-PVGPYIKYNLSPNKKVLLIMAGHGV---VLKNVKELCENLVKD 228
Db 247 RVFGLPVSFPRTILNKELRKELEIDNLPAVLLMGEGGEGMGPVOKTALALGDSLYNS 306
QY 229 -DOV---QVVVCGKNTALKESLSALEANGDKLVGYVERIDELFRITDCMITKPGGI 284
Db 307 KSNPTGQLLVICGRNKKVLASTASHEWK--IPVKGFGFETQEKMGWAGDCDIITAGPG 364
QY 285 TTEATAIGVPVILYKVPQGEKENANFFEDRGAIVVNRHIELESVTSLLA-DEDTLH 343
Db 365 TIAEALICGLPIILNDYIPQGEKGNVPYVDNGAGVTRSPKETAKIVADWFSNNKEELK 424
QY 344 RMKKNKIDHLANSSEVILEDKESMMTAKOK 377
Db 425 KMSENA--LKLQPEAVF--DIVKDIHLHSQQQQ 454

RESULT 6
Q9FZL5 PRELIMINARY; PRT; 465 AA.
AC Q9FZL5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE MGDG SYNTHASE TYPE C.
GN MGDG.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT "The Multigenic Family of MGDG synthases.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047398; BAB12041.1; -
SQ SEQUENCE 465 AA; 52990 MW; 05E0157012E50A14 CRC64;

Query Match 16.1%; Score 317; DB 10; Length 465;
Best Local Similarity 25.6%; Pred. No. 1.5e-12;
Matches 101; Conservative 90; Mismatches 171; Indels 32; Gaps 16;

RESULT	8	
PP3115		
ID	PP3115	PRELIMINARY; PRT; 525 AA.
AC	P31115;	
DT	01-JUN-1998 (TREMREL. 06, Created)	
DT	01-JUN-1998 (TREMREL. 06, Last sequence update)	
DT	01-MAR-2001 (TREMREL. 16, Last annotation update)	
DE	MONOACTOXYLDIACYLGlycerol SYNTHASE PRECURSOR (EC 2.4.1.46)	
DE	(1,2-DIACYLGlycerol 3-BETA-GALACTOSYLTRANSFERASE).	
QS	Cucumis sativus (Cucumber).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;	
OC	Cucurbitales; Cucurbitaceae; Cucumini.	
OX	NCBI_TaxID=3659;	
RN	[1]	
RP	SEQUENCE FROM N. A., AND PARTIAL SEQUENCE.	
RC	STRAIN=AONAGAIIBAI;	
RX	MEDLINE=97144442; PubMed=8990209;	
RA	Shimojima M., Ohta H., Iwamatsu A., Masuda T., Shioi Y.,	
RA	Takanaiya K.-I.;	

Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolutionary origin.;
Proc. Natl. Acad. Sci. U.S.A. 94:333-337(1997).
-1- FUNCTION: CATALYZES THE FORMATION OF MONOGALACTOSYLDIACYLGLYCEROL (MGDG) WHICH IS A MAJOR STRUCTURAL LIPID OF THE CHLOROPLAST.
-1- CATALYTIC ACTIVITY: UDP-GALACTOSE + 1,2-DIACYLGLYCEROL = UDP + 3-BETA-D-GALACTOSYL-1,2-DIACYLGLYCEROL.
-1- PATHWAY: GLYCOSYLATION.
-1- SUBCELLULAR LOCATION: CHLOROPLAST.
-1- SIMILARITY: TO E. COLI AND B. SUBTILIS UDP-N-ACETYLGALACTOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGALACTOSAMINE TRANSFERASE
EMBL: U62622; AAC49624.1; -
Mendel; 12656; Cusca; 1856; 12656.
KW Transferase; Glycosyltransferase; Transit peptide; Chloroplast.
FT TRANSIT 1 103 CHLOROPLAST.
FT CHAIN 104 525 MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
SQ SEQUENCE 525 AA; 57862 MW; 74FE586082EC48BA CRC64;

Query Match 15.4%; Score 303.5; DB 10; Length 525;
Best Local Similarity 25.7%; Pred. No. 1.3e-11;
Matches 100; Conservative 77; Mismatches 179; Indels 33; Gaps 13;

QY 5 KRVLIITANYNGH---VOVAKTLYEQVRLGQHVTVSNLYOESNPVSEVTQYLYLKS 61
DB 136 KRVLIILMSDTGGGHRASAEAKAFAEFENNYQ-VFIDTLWTDHTPWFNQLPRSYNEL 194
QY 62 FSTGK-QYRFLFYIG-VDKYINRKNFYFKMGKRLGELVDEHQPDIINTFPM----IV 116
DB 195 VKHGTUWKMTYVTPKVIHQSFATSTFTIAREVAKGLMKYRPDIISVHPLMQHVPIR 254
QY 117 VPEYRRRTGTVPTFNVMTDF-CLHKIWHNVNDKYVATDYVKEKLEIGTGHPSNVKIT 175
DB 255 ILRSKGLLNKIVFT-FVVTDLSTCHTPEFKLVTRCYCPSTEVAKRALTAGLQPSKLKF 313
QY 176 GPIRQPFESMPVGP---IYKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKDD 229
DB 314 GLPVRPSFKV--PIRKIELRKLGDENLPAVLMMGGEGMGPIATAKALSALYDEN 371
QY 230 Q-----QVVVCGKNTALKESLSALEANGDKLVGYVERIDELFRITDCMITKPGGIT 285
DB 372 HGEPIGOVLVICHGHNKLAGRLSIDWK--VPVQVKGFTVKMEECMGACDCIITKAGPGT 429
QY 286 LFEATAGVPIVLYKVPVQGEKENANFFEDRGAIVVNRHEEILLESVTSLLADE-DTLHR 344
DB 430 IAEAMIRGLPIILNDYIAGQAGNVYVVEGCGKFSKPKKAIANIVAKWFGPKADELLI 489
QY 345 MKNKIDHLANSSEVI-----LEDILKE 368
DB 490 MSQNA--LRLARPDVFKIVHDLJHVLKQ 516

RESULT 9
Q9FZL4 PRELIMINARY; PRT; 530 AA.
AC Q9FZL4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE MGDG SYNTHASE TYPE A.
GN GMDG A.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai K., Takamiya K., Ohta H.;
RT *cDNA cloning of MGDG synthase from tobacco and soybean.*;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047475; BAB11979.1; -

SQ SEQUENCE 530 AA; 57839 MW; D33C37FD53E90218 CRC64;

Query Match 15.3%; Score 302; DB 10; Length 530;
Best Local Similarity 27.0%; Pred. No. 1.6e-11;
Matches 104; Conservative 79; Mismatches 168; Indels 34; Gaps 17;

QY 5 KRVLIITANYNGH---VOVAKTLYEQVRLGQHVTVSNLYOESNPVSEVTQYLYLKS 61
DB 139 KRVLIILMSDTGGGHRASAEAKAFAEQERDDYQ-VFVTDLWADHTPWFNQLPRSY--S 195
QY 62 FSTGK-QYRFLFYIG-VDKYINRKNFYFKMGKRLGELVDEHQPDIINTFPMI--VV 117
DB 196 FLVKGHPLMKMTYGTAPRVVHQSFNEAATGTFIAREVAKGLMKYQPDIIISVHPLMQHVP 255
QY 118 PEVRRRTG--RVPTFNVMTDF-CLHKIWHNVNDKYVATDYVKEKLEIGTHPSNVKI 174
DB 256 LRLRSKGLLNKIVFTVTTITDLSCHTPTWFKLVTRCYCPTTDAVQAKLQSQSIKI 315
QY 175 TGIPIRQPFESMPVGP---IYKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKDD 228
DB 316 FGLPVRPSFKV--PVQPKDELRELGMDEDLPAVLMMGGEGMGPIETARALGDSLY-D 372
QY 229 DOV-----QVVVCGKNTALKESLSALEANGDKLVGYVERIDELFRITDCMITKPGG 283
DB 373 ENIGAPVGOILVICGRNKKLANKLSSINWK--VPVQVKGFTVKMEECMGACDCIITKAGP 430
QY 284 IYLTEATAGVPIVLYKVPVQGEKENANFFEDRGAIVVNRHEEILLESVTSLLADE-TL 342
DB 431 GTIAEAGIRGLPIILNDYIAGQAGNVYVVEGCGKFSKPKKAIKIVAEVFGPKRAYEL 490
QY 343 HRMKNKIDHLANSSEV--ILEDI 365
DB 491 QMSQNA--LRLARPDVFKIVHDL 513

RESULT 10
Q9MU68 PRELIMINARY; PRT; 533 AA.
AC Q9MU68;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
GN MGD1.
OS Arabidopsis thaliana (Mouse-ear cress).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Jarvis P., Doermann P., Peto C.A., Lutes J., Benning C., Chory J.;
RT *Galactolipid-Deficiency and Abnormal Chloroplast Development in the Arabidopsis MGD Synthase 1 Mutant.*;
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
DR EMBL; AF241797; AAF65066.1; -
KW Chloroplast.
SQ SEQUENCE 533 AA; 58537 MW; E581E67317CB9CC8 CRC64;

Query Match 15.3%; Score 301.5; DB 8; Length 533;
Best Local Similarity 24.8%; Pred. No. 1.7e-11;
Matches 101; Conservative 86; Mismatches 172; Indels 49; Gaps 17;

QY 5 KRVLIITANYNGHVOVAKTLYEQVRLGQ-----HVTVSNLYOESN--PIVSEVTQ 55
DB 142 KRVLIILMSDTGGGHRASA-----EAIARAANQEFGEYQVFTITDLWTDHTPWFNQLPRS 196
QY 56 YLYLKSFSIGKQFYRLFYGVVD-KYNNKKNFYFKMGKRLGELVDEHQPDIINTFPM 114
DB 114 YLYLKSFSIGKQFYRLFYGVVD-KYNNKKNFYFKMGKRLGELVDEHQPDIINTFPM 114

Db 197 YNFKVKG---TLKMTYIGTSPRIVHQSFAATSTFIARETAOGLMKYQPDIIISVHPL 253
QY 115 I--VPEYRRRTG---RVIPTFNVMTDF--CLHKIWVHENVKYYVATDYVKEKLEIGTHP 169
Db 254 MOHVPLRVLSKGLLKIVFTTITDSTCHPTWFKLVTRCYCPSTEVAKRAQAGLET 313
QY 170 SNVKITGIPRQPEESMPVGP---IYKYNLSPNKKVLLIMAGAHGV--LKNVKELCEN 224
Db 314 SQIKYGLPVRPSFVK---PVRPKVELRGLMDENLPAVLLMGEGMGPIEATARALAD 371
QY 225 LVKDDQV-----QVVVCGKNTALKESALEAENGDKLVGLYVERIDELFRITDCMIT 279
Db 372 ALYDKNLGEAVGVLLICGRNKKLSKLSLDWK--IPVQVKGFTKMEECMGACDCIIT 429
QY 280 KPGGITTLEATAIGPVILYKPVQGEKENANFFEDRGAALVNRHHEILSVTSLLAD- 338
Db 430 KAGPOTIAEAMIRGLPIILNGYIAGQAGNVPYVWNC---GKFSKSPKEISKIVADW 485
QY 339 ---EDTLHRMKNKIDHLANSSEVILEDLKESMMTAKQAKVLS 382
Db 486 FGPASKELEIMSONA--LRLA-KPEAVFKIVHDMHELVRKKNLSPLQS 530
RESULT 11
O81770 PRELIMINARY; PRT; 533 AA.
AC O81770;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MONOACTOSYLDIACVGLYCEROL SYNTHASE - LIKE PROTEIN
DE (MONOACTOSYLDIACVGLYCEROL SYNTHASE-LIKE PROTEIN) (MGDG SYNTHASE
DE TYPE A).
GN F28M20.30 OR AT4G31780 OR MGDA.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Hobeisel J.,
RA Mewes H.W., Mayer K.F.X., Schueller C., Bevan M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
RA Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Terryn N., Ardiles W., Buysschaert C., Dasseville R., De Clerck R.,
RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RA Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT "The Multigenic Family of MGDG synthases";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031004; CA19745.1; -
DR EMBL; AL161579; CAB79896.1; -
DR EMBL; AB047399; BAB12042.1; -
DR Mendel; 32401; Arath;1856;32401.
SQ SEQUENCE 533 AA; 58537 MW; E581E67317CB9CC8 CRC64;

Query Match 15.3%; Score 301.5; DB 10; Length 533;
Best Local Similarity 24.8%; Pred. No. 1.7e-11;

Matches 101; Conservative 86; Mismatches 172; Indels 49; Gaps 17;
QY 5 KRVLLITANYGNHGVAKLYEQVRLGFQ-----HVTVSNLYQESN--PIVSEVTQ 55
Db 142 KKVLLMSDGTGGGRASA-----EAIKAFNQEFGEYQVFTDWTHTPWPENQLPRS 196
QY 56 YLKSFSGIKQFVRLFYGVYD-KIYNKRKFNIYFKMGKRLGELVDEHQPDIILTFPM 114
Db 197 YNFKVKG---TLKMTYIGTSPRIVHQSFAATSTFIARETAOGLMKYQPDIIISVHPL 253
QY 115 I--VPEYRRRTG---RVIPTFNVMTDF--CLHKIWVHENVKYYVATDYVKEKLEIGTHP 169
Db 254 MOHVPLRVLSKGLLKIVFTTITDSTCHPTWFKLVTRCYCPSTEVAKRAQAGLET 313
QY 170 SNVKITGIPRQPEESMPVGP---IYKYNLSPNKKVLLIMAGAHGV--LKNVKELCEN 224
Db 314 SQIKYGLPVRPSFVK---PVRPKVELRGLMDENLPAVLLMGEGMGPIEATARALAD 371
QY 225 LVKDDQV-----QVVVCGKNTALKESALEAENGDKLVGLYVERIDELFRITDCMIT 279
Db 372 ALYDKNLGEAVGVLLICGRNKKLSKLSLDWK--IPVQVKGFTKMEECMGACDCIIT 429
QY 280 KPGGITTLEATAIGPVILYKPVQGEKENANFFEDRGAALVNRHHEILSVTSLLAD- 338
Db 430 KAGPOTIAEAMIRGLPIILNGYIAGQAGNVPYVWNC---GKFSKSPKEISKIVADW 485
QY 339 ---EDTLHRMKNKIDHLANSSEVILEDLKESMMTAKQAKVLS 382
Db 486 FGPASKELEIMSONA--LRLA-KPEAVFKIVHDMHELVRKKNLSPLQS 530
RESULT 12
O9FZL3 PRELIMINARY; PRT; 535 AA.
AC O9FZL3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MGDG SYNTHASE TYPE A.
GN NTMGD A.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
[1]
RN SEQUENCE FROM N.A.
RA Awai K., Takamiya K., Ohta H.;
RT "cDNA cloning of MGDG synthase from tobacco and soybean";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047476; BAB11980.1; -
SQ SEQUENCE 535 AA; 59589 MW; B1B2067E86EDE477 CRC64;
Query Match 15.0%; Score 295; DB 10; Length 535;
Best Local Similarity 26.2%; Pred. No. 4.5e-11;
Matches 102; Conservative 76; Mismatches 169; Indels 42; Gaps 15;
QY 5 KRVLLITANYGNHGVAKLYEQVRLGFQ-----HVTVSNLYQESNPIVSEVTQYL 57
Db 145 KKVLLMSDGTGGGRASA-----EAIKAFNQEFGEYQVFTDWTHTPWPENQLPRS 199
QY 58 YLKSFSGIKQFVRLFYGVYD-KIYNKRKFNIYFKMGKRLGELVDEHQPDIILTFPMI-- 115
Db 200 YNFKVHGSILRWYTYATAPRLVHQTNPATSTFIAREVAKGLMKYQPDIIISVHPLMOH 259
QY 116 VPEYRRRTG---RVIPTFNVMTDF--CLHKIWVHENVKYYVATDYVKEKLEIGTHPSN 171
Db 260 VPLRLRSKGLLKIIIFT-TVITDSTCHPTWFKLVTRCYCPSEVAKRALRAGLKPQY 318
QY 172 VKITGIPRQPEESMPVGP---IYKYNLSPNKKVLLIMAGAHGV--LKNVKELCENL 225
Db 319 LKVIYGLPVRPSFVK---PVPKVELRGLMDENLPAVLLMGEGMGPIEATARALGDAL 376

```

Query Match      14.9%; Score 293; DB 10; Length 468;
Best Local Similarity 25.1%; Pred. No. 5e-11;
Matches 98; Conservative 86; Mismatches 172; Indels 34; Gaps 14;

QY      5  KRVLILTANTYNGNHVQVAKTLYEOC-VRLGQOH-VTVSNLYQESN--PIVSEVTQYLXLK 60
      | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      69  KNVLILMSDTGGGHRASAEAIRDAFKIEFGDKYRVIVKDVWVKELTGTPLNDMERSYKPMV 128
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      61  SFSIGKGFYRLFFYG-----VDKTYNKRKKNFYKMGKNKRLGELVDEHQPDIIITPFM- 114
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      129  KH---VQLWKVAFHSTSPKWIHSCYLAALAAIYAKVEAGL---MEYKPEIILSVHPLM 181
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      115  ----IVVPEYRRRTGRVIPFNVTDF-CLHKIKVHNVDKYYVATDVYKELKLEIGTHP 169
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      182  QH1PLWLWKQWELQKRLV-FVTVITDITLNTCHPTWFHPGVNRCYCPSQEVAKRALFDGLDE 240
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      170  SNVKITGIPTRPOEESMPV-GPIYKKNLSPNKKVLLIMAGHGV--LKNVKELCENLV 226
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      241  SQVRVFLGPRVPSFARAVLVKDDLKELEMDODDLRAVLMLGGGGMGSPVKETAKALEEFL 300
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      227  KDDQ-----VOVVVVCNKNTALKESLSALEAENGDKLKVLYGVVERIDELFRITDCMITKP 281
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      301  YDKNRKRPIGOMVVICGRNKKLASALEAIDWK--IPVKVGFETOMEKWMGACDCIITKA 358
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      282  GGITLTATAIGVPVILYKVPVGOEKENANFFEDRGAIVVNRHBEILESVTSLLADE-D 340
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      359  GPGTIAESLIRSLPIILINDIPGOEKNVPVYVWENGAGVTRSPRKTARIIVGEWFSTKD 418
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      341  TLHRMKNKINRDLHANSSEVILEDKKESE 370

```

[illegible]

RESULT 2
US-08-429-054A-11
; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN; VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHETASE (SPS),
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; TITLE OF INVENTION: CDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,054A
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 842,337
; FILING DATE: 20-March-1992
; APPLICATION NUMBER: PCT/FR 91/00593
; FILING DATE: 18-July-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: French 90402094.9
; FILING DATE: 20-July-1990
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles A. Musserlian
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146.1137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
US-08-429-054A-11

Query Match 5.7%; Score 112.5; DB 2; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.012;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;
QY 125 GRVPIPTFNMV---TDFCLHKIWHENVVDKYVATDYVVKELLEI-GTHPSNVKITGIPR 180
DB 427 GRYPMRVWVIPPGMDF--SNVVVHEDID---GDGVDKDDIVGLEGASPKSMP----- 473
QY 181 PQFEESM-----PVGPIYKYNLSNKKVLLIMAGAHGVLKNVKELCNLVKDDQVQVVV 235
DB 474 PIWAEVWFELTNPHKPMILALS RDPDKNITTLVKAFCGRPLREL-----ANLTL 524
QY 236 VCGKNWTKALSLSALBAENGDKL-KVLGYVERID-----ELFRITDCM 277
DB 525 IMGN----RDDIDMSAGNASVLTTLKLDIKDYLYGSVAFFPKHHNQADVPEIYRLAAKM 580
QY 278 -----ITKPGGITITATAIGVPVILYKVPVQGEKENANFFEDRGAIVVNRHEE--- 327

DB 581 KGVFINPALVEPFGTLTIEAAHGLPIVATKNGGPDVITNA-----LNNGLLVDPHDQNA 635
QY 328 ILSVTSLLADEDTLHRMKN-KDKLHL 354
DB 636 IADALLKLVADKNLWQECRRNGLRNIHL 663
RESULT 3
US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
; GENERAL INFORMATION:
; APPLICANT: Van Assche, C.
; APPLICANT: Lando, D.
; APPLICANT: Bruneau, J. M.
; APPLICANT: Voelker, T.
; APPLICANT: Gervais, M.
; TITLE OF INVENTION: MODIFICATION OF SUCROSE
; TITLE OF INVENTION: PHOSPHATE
; TITLE OF INVENTION: SYNTHASE IN PLANTS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Barbara Rae-Venter
; STREET: 260 Sheridan Avenue, Suite 440
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,777
; FILING DATE: NOT YET ASSIGNED
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,471
; FILING DATE: 27-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara Rae-Venter
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE.072.0205
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)328-4400
; TELEFAX: (415)328-4477
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-777-7

Query Match 5.7%; Score 112.5; DB 2; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.012;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;
QY 125 GRVPIPTFNMV---TDFCLHKIWHENVVDKYVATDYVVKELLEI-GTHPSNVKITGIPR 180
DB 427 GRYPMRVWVIPPGMDF--SNVVVHEDID---GDGVDKDDIVGLEGASPKSMP----- 473
QY 181 PQFEESM-----PVGPIYKYNLSNKKVLLIMAGAHGVLKNVKELCNLVKDDQVQVVV 235
DB 474 PIWAEVWFELTNPHKPMILALS RDPDKNITTLVKAFCGRPLREL-----ANLTL 524
QY 236 VCGKNWTKALSLSALBAENGDKL-KVLGYVERID-----ELFRITDCM 277
DB 525 IMGN----RDDIDMSAGNASVLTTLKLDIKDYLYGSVAFFPKHHNQADVPEIYRLAAKM 580

QY	278	-----ITKFGGTLT	TAATG	VPVILYK	VPQGEKENAN	FFEDRGAAI	VVVRHEE	---	327
Db	581	KGVF	INP	ALVEP	FGTLT	IAAAHGL	PIVATK	GGPVDITNA	-----LNNGLLVDPHDQNA
QY	328	ILES	VTSL	LADE	DTL	HRMKN	-IK	LIHL	354
Db	636	IAD	ALL	KL	VAD	KNL	WQEC	RNRGLNIHL	663

RESULT 4
US-09-051-341-7
; Sequence 7, Application US/09051341
; Patent No. 6124528
; GENERAL INFORMATION:
; APPLICANT: Shewmaker, C. K.
; TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
; NUMBER OF SEQUENCES: 14

```
Query Match.          5.7%; Score 112.5; DB 4; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.012;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;
```

QY	125	GRVIPTFNMV--TDFCFLHKIWHENVNDKYYKVATDYVVEKKLLEI-GTHPSNNVKITGIPR	180
		: : :: : : :	
Ddb	427	GRYPMPWVIPPGMDF--SNVVYHEDID---GDGVKDVIDGLEGASPKSM-----	473
QY	181	POFEESM-----PVGPVIKYKNISPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVV	235
		: : : : : : : : : : : :	
Ddb	474	PIAWEMRELTNPHKPMILALSRDPDKKNITTLVKAFCRCPLREL-----ANLTL	524
QY	236	VCGKNTALKESLSALEAENGDKL-KVLGYVERID-----ELFRITOCM	277
		: : : : : : : : : : : :	
Ddb	525	TGN-----RDDIDMSAGNASVITVLKLKDYDLKVDYSVAFFPKHHNQADPVPYILAAM	580

[illegible]

```

1  RESULT      5
2  PCT-US92-00282-7
3  ; Sequence 7, Application PC/TUS9200282
4  ; GENERAL INFORMATION:
5  ; APPLICANT: OWENS, IDA S.
6  ; APPLICANT: RITTER, JOSEPH K.
7  ; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
8  ; TITLE OF INVENTION: THEREIN.
9  ; NUMBER OF SEQUENCES: 40
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
12 ; STREET: 1615 L STREET, N.W.
13 ; CITY: WASHINGTON
14 ; STATE: D.C.
15 ; COUNTRY: U.S.A.
16 ; ZIP: 20036-5601
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Floppy disk
19 ; COMPUTER: IBM PC compatible
20 ; OPERATING SYSTEM: PC-DOS/MS-DOS
21 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: PCT/US92/00282
24 ; FILING DATE: 19920110
25 ; CLASSIFICATION: 435
26 ; ATTORNEY/AGENT INFORMATION:
27 ; NAME: SCOTT, WATSON T.
28 ; REGISTRATION NUMBER: 26581
29 ; REFERENCE/DOCKET NUMBER: 91532-PCT
30 ; TELECOMMUNICATION INFORMATION:
31 ; TELEPHONE: 202-861-3000
32 ; TELEFAX: 202-822-0944
33 ; INFORMATION FOR SEQ ID NO: 7:
34 ; SEQUENCE CHARACTERISTICS:
35 ; LENGTH: 529 amino acids
36 ; TYPE: AMINO ACID
37 ; STRANDEDNESS: single
38 ; TOPOLOGY: linear
39 ; MOLECULE TYPE: protein
40 ; PCT-US92-00282-7

```

Query Match	5.6%	Score 110;	DB 5;	Length 529;
Best Local Similarity	20.5%;	Pred. No. 0.0071;		
Matches	79;	Conservative 57;	Mismatches 137;	Indels 112; Gaps 16;
QY	40	SNLYQESNPITVSEWTOY-----LYLK-----SFSIGKQFYRLF-----	72	
DB	95	NNHFAASSPLMAPLREYRNMMIVIDCMFCFCSQSLKDSATLSFLRENOQDFALFTDPAMP	154	
QY	73	-----YGVGDKIYNKKRKNFYKMGKNRIGELVDEHQPDIIINTFPNIVVPEYRRTR	126	
DB	155	GVILAEYLKPSIYLFPG-----FPCSLHEIGQ-----SPSPSVVVPREYTKFSD	199	
QY	127	VIPTE-----NVMYDFCLHKIWHVENVDKYV-ATDYVKEKLEIGTHPSNVKITG	176	
DB	200	HM-TFFCRLANFANTILENYLHCLY-----SKYEILASDLLKRDVSLPALHQNSLWLR	253	
QY	177	IPTRQFESMPVGP-----IYKYNLSPNKKVLLIMAGAHGYL-----	215	
DB	254	YDF--VFYEYPRVPYNNMIFITGGTNCKKKGNLSOFEFAYVNASGEHGYVWFSGLSMVSEIP	311	

; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D, Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: Case No. 2580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-420-3300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US91-07035-12

Query Match 5.5%; Score 108; DB 5; Length 892;
Best Local Similarity 19.4%; Pred. No. 0.026;
Matches 72; Conservative 63; Mismatches 135; Indels 102; Gaps 15;

QY 32 LGFQHVTVSNLYQSNPIVSEVTOYLK-----SFSIGKGFYRLFYGVVDKIYKRNKN 86
DB 431 LGYKMSIDFELNVNPLFGNDFSYVPLRAVEYSCEDADVYRIF----- 476
QY 87 IYFKMGKRLGELVDEHOPDIITFTPMVVPYRRRTGRVPTFNVMTDFCLHKIWVHE 146
DB 477 -----RKLGRKIYENEMKLFIEIEM-----PLIDVSEMLNGVYFDE 515
QY 147 NV-----DKYVATDYVKEKLELTGTHPSNVKINGIPRPFESMPVGPI-YKKYNLSP 200
DB 516 EYLKELSKYQKMDGIKEKVFETAGETFNL-----NSSTQVAYILPEKLNIA 564
QY 201 NKKVLLIMAGHGVKLVKNVLCENLVDDQVQVGVVCGKNTALKEISALEAENGDKLV 260
DB 565 YKKT-----ATGKFSTNAEYLELSKEHEIAKLL-----LEYRKYQKLKS 604
QY 261 LGYVE-----RIDELFRITDCMITKPGGITLTETAIGVPVILYKVPVGQKEN 309
DB 605 T-YIDSLPLSNKTNRVHTTFHQTG---TSTGRLSSNPNLQNLPT---RSEEGKEIRK 657
QY 310 ANFFEDGAAIVVNRHEEI-LESVTSLLADEDTLHRMKKNIKOLHLANSSEVILEDLKE 368
DB 658 AVRPRQDWWTLGADYSQIELRVLAHVSKDENLILKAFKEDL-DIHTTAAK-----IFGV 711
QY 369 SEMTAKOKAKV 380
DB 712 SEMFVSEOMRRV 723

RESULT 9
US-08-843-530B-36
; Sequence 36, Application US/08843530B
; Patent No. 5939306
; GENERAL INFORMATION:
; APPLICANT: Selitrennikoff, Claude
; APPLICANT: Agnan, Jacqueline
; APPLICANT: Alex, Lisa A.
; APPLICANT: Simon, Melvin I.
; TITLE OF INVENTION: Osmosensing Histidine Kinases
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco

; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/843,530B
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MacKnight, Kamrin T.
; REGISTRATION NUMBER: 38,230
; REFERENCE/DOCKET NUMBER: UTC-02717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1220 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-843-530B-36

Query Match 5.4%; Score 105.5; DB 2; Length 1220;
Best Local Similarity 20.2%; Pred. No. 0.076;
Matches 76; Conservative 60; Mismatches 124; Indels 117; Gaps 19;

QY 11 TANYGN---GHVOVAKTYEQVRLGFQHVTVSNLYQES-----NPVSEVTOYL--- 57
DB 47 TSNKNLRSRLYTAOLKSSQI-----DOTNLVLYQAYTLASRDALQSSLTSYVAGNK 101
QY 58 ----YLSFSISIGKOFY---RLFYGVVDKIYKRNKNFYFKMGRKRLGELVDEHOPDIIN 110
DB 102 SADNWDVSLSVIQKFLSSSNLFY--VAKVYDS-SFNAVLTNNTNGTGLIPE---DVLDS 155
QY 111 TFMVIVPEYRRRTGRVIP---TFNVMTDFCLHKIWHNVNDKYVYATDYVKEKLEIG 166
DB 156 LFPL-----STDTPLSPSLETIGILTDPVLN-----STDYLSMSLPIF 194
QY 167 THPS-----NVKITGPIRPQEE-----ESMPYGPYKYNLSPNKKVLLIMA 209
DB 195 ANPSIIITDSRVGYITIIISAEGLKSVFNDTALHSTTALISAVTNSOCKASGYHVF 254
QY 210 GAHGVLANVRELCELVKDDQVQVGVVCGKNTALKEISALEAENGDKLVGYVERIDE 269
DB 255 PPGYRSDDLQKVFESIKNDJTFISSAFRNGKGSLLKQT-NILSTRN---TALGY----- 303
QY 270 LFRITDC-----MITKPGGITLTETAT-----AIGVPVIL-----YK 300
DB 304 ----SPCSFNLVNWVAIVSQPESVFLSPATKLAKIITGTVIAIGVFILLTPLAHWAVQ 359
QY 301 PVPQOEKENANFFEDRG 317
DB 360 PIVRLQKATELITEGRG 376

RESULT 10
US-09-007-476-2
; Sequence 2, Application US/09007476
; Patent No. 6159949
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6159949el Ftsy
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

Query Match
5.18; Score 101; DB 2; Length 403;

Best Local Similarity 20.8%; Pred. No. 0.04;
Matches 61; Conservative 50; Mismatches 110; Indels 72; Gaps 12;
QY 126 RVIPFNWTFCLKIWHNVNDKYVATDVYKELK-----EIGT 167
Db 23 RPIPSFDDMP-----LHQNLLRGIYSYGFEKPSIIQORAIAPFTRGGDIIAQASGT 74
QY 168 HPSNVKITGIPRPOFESMPVGPYKYNLSPNKKVLL-----IMAGAHGVLNKVKELCE 223
Db 75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNKAFC 129
QY 224 NL-----VKDD-----QVQVVVVC-----KNTALK-ESLSALEAENGDKLVGLVYV 264
Db 130 TFVGGRVODDLRKLQAGVWVAVGTPGRVSDVIKRCALRTESLRVLVLDDEADMLSQGFA 189
QY 265 ERIDELFRITDCMITYKPGGIIITTEATAIGVPIVLYKVPVQGEKENANFFEDRGAIVNVR 324
Db 190 DQIYEIER-----FLPKDIOQVAFSAT-----MPEEVLTKKFMRPDPVRILVKR 234
QY 325 HEEILSVTSLADEDTLHRMKNKIKDLH--LANSSEVILEDKESMMTAK 375
Db 235 ESITLEGIKQFFIAVEEHL-DTLMDLVETVSIQSVIFANTRRKVDWIAEK 286

RESULT 15
US-08-634-642-2
; Sequence 2, Application US/08634642
; Patent No. 5879687
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF
; TITLE OF INVENTION: PROTECTIVE IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,642
; FILING DATE: 18-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210134.404C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-634-642-2

Query Match 5.1%; Score 101; DB 2; Length 403;
Best Local Similarity 20.8%; Pred. No. 0.04;
Matches 61; Conservative 50; Mismatches 110; Indels 72; Gaps 12;
QY 126 RVIPFNWTFCLKIWHNVNDKYVATDVYKELK-----EIGT 167
Db 23 RPIPSFDDMP-----LHQNLLRGIYSYGFEKPSIIQORAIAPFTRGGDIIAQASGT 74

QY 168 HPSNVKITGIPRPOFESMPVGPYKYNLSPNKKVLL-----IMAGAHGVLNKVKELCE 223
Db 75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNKAFC 129
QY 224 NL-----VKDD-----QVQVVVVC-----KNTALK-ESLSALEAENGDKLVGLVYV 264
Db 130 TFVGGRVODDLRKLQAGVWVAVGTPGRVSDVIKRCALRTESLRVLVLDDEADMLSQGFA 189
QY 265 ERIDELFRITDCMITYKPGGIIITTEATAIGVPIVLYKVPVQGEKENANFFEDRGAIVNVR 324
Db 190 DQIYEIER-----FLPKDIOQVAFSAT-----MPEEVLTKKFMRPDPVRILVKR 234
QY 325 HEEILSVTSLADEDTLHRMKNKIKDLH--LANSSEVILEDKESMMTAK 375
Db 235 ESITLEGIKQFFIAVEEHL-DTLMDLVETVSIQSVIFANTRRKVDWIAEK 286

Search completed: June 29, 2001, 08:59:39
Job time: 177 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 11:53:36 ; Search time 2892.01 Seconds
(without alignments)
5214.731 Million cell updates/sec

Title: US-09-668-788-3
Perfect score: 975
Sequence: 1 atgtttactcaataaaaa.....gtaaaacgtgatactccag 975

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_rod.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_vil2.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_rod.*
95: gb_rod2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	975	100.0	975	9	AX016297	Sequence
2	964	98.9	7791	3	SAY14370	Y14370 Staphylococ
3	636.2	65.3	3118	2	AF270166	AF270166 Staphyloc
4	636.2	65.3	3493	2	AF270394	AF270394 Staphyloc
5	437	44.8	2600	9	AR106079	AR106079 Sequence
6	437	44.8	2600	10	E35634	E35634 MurE. 2/200
7	118	12.1	1149	9	AX016296	AX016296 Sequence
8	118	12.1	35040	2	BACYACA	L77246 Bacillus su

9 118 12.1 213680 2 BSUB0012 Z99115 Bacillus su
10 72 7.4 242893 83 CEY53C12 Z92859 Caenorhabdi
11 60.2 6.2 1496 6 CEY53C12D AL0331231 Caenorhab
12 57.4 5.9 1648 13 ATMGDGSYN AJ000331 Arabidops
13 56.8 5.8 44352 6 CELH28G03 AF098501 Caenorhab
14 56.4 5.8 152878 6 CEYL8D10A AL034393 Caenorhab
15 56.4 5.8 247332 83 CEYL8D10 AL008871 Caenorhab
16 55.4 5.7 148061 87 AC010727 AL007027 Homo sapi
17 55 5.6 1574 12 AB047398 AB047398 Arabidops
18 54.6 5.6 1141 10 AX083744 AX083744 Sequence
19 54.4 5.6 175415 82 AL590063 AL590063 Homo sapi
20 54.2 5.6 169437 81 AL391538 AL391538 Homo sapi
21 54 5.5 214631 72 AC060805 AC060805 Homo sapi
22 53.6 5.5 53932 68 AC023371 AC023371 Homo sapi
23 52.8 5.4 54345 4 AC084152 AC084152 Caenorhab
24 52.6 5.4 1154 5 AF044864 AF044864 Pteris ra
25 52.6 5.4 1189 80 AL359196 AL359196 Homo sapi
26 52.4 5.4 867 53 CNS075BG AL429890 clone BA0
27 52 5.3 205429 60 AC005506 AC005506 Plasmodi
28 51.4 5.3 129388 69 AC025184 AC025184 Homo sapi
29 51.2 5.3 1775 12 AB047399 AB047399 Arabidops
30 51.2 5.3 1929 13 AF241797 AF241797 Arabidops
31 51 5.2 1038 53 CNS06L7M AL403832 T3 end of
32 51 5.2 186625 86 AC006487 AC006487 Homo sapi
33 50.8 5.2 155317 73 AC067973 AC067973 Homo sapi
34 50.4 5.2 100925 87 AC012627 AC012627 Homo sapi
35 50.4 5.2 170432 67 AC022145 AC022145 Homo sapi
36 50.4 5.2 235532 87 AC008739 AC008739 Homo sapi
37 50.2 5.1 910 53 CNS01GHP AL42826 Anopheles
38 50.2 5.1 156060 60 AC004153 AC004153 Plasmodi
39 50 5.1 1496 6 CEY53C12D AL033123 Caenorhab
40 50 5.1 149752 85 AC004616 AC004616 Homo sapi
41 50 5.1 242893 83 CEY53C12 Z92859 Caenorhabdi
42 49.8 5.1 759 53 CNS06QXV AL411257 T7 end of
43 49.8 5.1 1013 53 CNS06RQV AL412260 T7 end of
44 49.8 5.1 166331 80 AL360177 AL360177 Homo sapi
45 49.8 5.1 171863 91 AP000459 AP000459 Homo sapi

ALIGNMENTS

RESULT 1
AX016297 975 bp DNA PAT 07-SEP-2000
LOCUS Sequence 2 from Patent WO9949052.
DEFINITION AX016297
ACCESSION AX016297
VERSION AX016297.1 GI:10041860
KEYWORDS Staphylococcus aureus.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus
REFERENCE Staphylococcus aureus
AUTHORS Bacteria; Firmicutes; Bacillus/Clostridium group;
TITLE Bacillus/Staphylococcus group; Staphylococcus.
JOURNAL 1 (bases 1 to 975)
Zaeheringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.
Processive glycosyltransferase
Patent: WO 9949052-A 2 30-SEP-1999;
FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
(DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
UND VERWERB (DE)

FEATURES
source Location/Qualifiers
1..975
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
BASE COUNT 350 a 152 c 169 g 304 t
ORIGIN

Query Match 100.0%; Score 975; DB 9; Length 975;
Best Local Similarity 100.0%; Pred. No. 1.5e-164;
Matches 975; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atggttactcaaaataaaagattgatttactggtcattcggttaacggtcatatg 60

Db 1 ATGGTACTCAAAATAAAGATATATGATTATTAAGTGGCTCAATTCGGTAACGGTCATATG 60
Qy 61 caagtacacagagtatcgttaatacaactaatgatataatgaatctacagaccattagcgtc 120
Db 61 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTAAGCGTC 120
Qy 121 attgagcacagatttattatgaagctcacaatttgacttcttatttgtaaaaaatgg 180
Db 121 ATTGACACAGATTATTTATGAAGCTCATCAATTTTGACCTCTCTATTGTAANAATGG 180
Qy 181 tatataagctgtttaataattttagaataatgtacaaaggggtttttattacagccgcccc 240
Db 181 TATATCAATAGCTTTAAATATTTAGAAATATGTACAAAAGGGTTTTATTACAGCGCCCA 240
Qy 241 gataactagacaaatgtttttacaaataactatgaacttaataagttaattattatg 300
Db 241 GATAAACTAGACAAATGTTTTACAATACTATGGACTTAATAAAGTAAATTAATTTATG 300
Qy 301 ataaaagaaagccagatttaattattataacgttttcacaccaggtatgtcggtaacta 360
Db 301 ATAAAAGAAAAGCCAGATTTAATATTATTAAACGTTTCTTACACAGCTATGTCGGTACTA 360
Qy 361 actgagcaatttaacatttaattatccagttgctacagtgatgacagactatcgtctacat 420
Db 361 ACTGAGCAATTTAACATTAAATATTCAGTTGCTTACAGTGTACAGACTATCGCTTACAT 420
Qy 421 aaaaactgattacgctgatttcaacaagatatattgtggcaacaaagaaacaaacaa 480
Db 421 AAAAATGGATTACGGCGTATTCAACAAGTATTATGTGGCAACAANAAGAAACAA 480
Qy 481 gacttcagacgtaggtattgcttcctcaacagtttaagtgacaggtattcctattgat 540
Db 481 GACTTCATAGACGTAGTATTGATCTCTCAACAGTTTAAAGTACAGAGTATTCCTATTGAT 540
Qy 541 acaaatattgaacgcctatttaatacaaaagcagttggttaataagaacaacttagatcca 600
Db 541 AACAAATTTGAAACGGCTCTTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCA 600
Qy 601 gataagcaaaactattttaatgtcagctggtgcatgttggtggtatctctaaaggtttgcacg 660
Db 601 GATAAGCAAACTATTTTAATGTGAGTGGTGGTGTATCTAAAGGTTTTGACACG 660
Qy 661 atgattactatatttagcgaaaagtgcataatgcacaaagtagtattgattggttaag 720
Db 661 ATGATTACTATATATAGCGAAAAGTGCAAATGCAAATGCAAATGATGATTGTTGGTAA 720
Qy 721 agcaagagctaaagcgtttcttaacagctaaagttaaataacagagaaatgtattgatt 780
Db 721 AGCAAGAGCTAAAGCGTTCTTTAACAGCTAAGTTTAAATTAACGAGAAATGTATTGATT 780
Qy 781 ctagggtataccaaacacatgaatgaatggatggatcgaactcaactatgattacgaaa 840
Db 781 CTAGGTTATACCAACACATGAATGAATGGATGGATGCATCAAGTCAACTTATGATTACGAAA 840
Qy 841 cctggtgattcacataaactgaaggttcgccgttgattcccaatgatttccctaaat 900
Db 841 CCTGGTGGTATCAAAATAACTGAAGGTTCCCGCGTTGTTATTCATCATTTTCCTAAAT 900
Qy 901 cctgcacctggtcgaagcgttgaaaatgccttttactttgaaagaaaaggttttgggttaa 960
Db 901 CCTGCACCTGGTCAAGAGCTGAAATGCCTTTTACTTTGAAGAAAAGGTTTTTGGTAAA 960
Qy 961 acgctgatactccag 975
Db 961 ACCTGATACTCCAG 975

RESULT 2
SAY14370 7791 bp DNA BCT 24-JUN-1998
LOCUS SAY14370
DEFINITION Staphylococcus aureus RF3, murE, yfpP genes.
ACCESSION Y14370

VERSION Y14370.1 GI:3256221
 KEYWORDS mure gene; peptide chain release factor 3; RF3 gene;
 SOURCE UDP-N-acetylmuramyl-tripeptide synthetase; ypfp gene.
 ORGANISM Staphylococcus aureus.
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Staphylococcus aureus group; Staphylococcus.
 REFERENCE 1 (bases 1 to 7791)
 AUTHORS Ludovice, A.M., Wu, S. and de Lencastre, H.
 TITLE Molecular cloning and DNA sequencing of the Staphylococcus aureus
 UDP-N-acetylmuramyl tripeptide synthetase (mure) gene, essential
 for the optimal expression of methicillin resistance
 JOURNAL Microb. Drug Res. 4, 85-90 (1998)
 REFERENCE 2 (bases 1 to 7791)
 AUTHORS Ludovice, A.M.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUL-1997) A.M. Ludovice, Instituto de Tecnologia
 QIomica e Biologica, Universidade Nova de Lisboa, R. da Quinta
 Grande 6, Apartado 127, 2780 Oeiras, PORTUGAL
 REMARK Revised by author 16-JAN-98
 COMMENT Related sequence: L77246.
 FEATURES
 source 1..7791
 /organism="Staphylococcus aureus"
 /strain="COL"
 /db_xref="taxon:1280"
 complement(1435..3114)
 /gene="RF3"
 complement(1435..3000)
 /gene="RF3"
 /codon_start=1
 /transl_table=11
 /product="peptide chain release factor 3"
 /protein_id="CAA74739.1"
 /db_xref="GI:3256222"
 /db_xref="SPTREMBL:O86490"
 /translation="MQLQVEERKTFALISHPDAGKTLTEKLLYFSGAIREAGTVK
 GKLLNLRQVTKMVEGERGISVTSVMQFDYDEINILDPGHEDESEDTYRLMA
 VDSAVMIDCAKGFPPPLTKLFKVCCKMRGIPFTFKLDKRVGKPPFELLIDETLMA
 IETYPNNPIMGQSFPGFIDIRKSKTIEPPEDEENILHNDQDFELEEDHAIITDSDFE
 QAIIEMLVEAGEAGDFNDALLSGDLTPVFGSALANFGVONFLNAYVDFAPMPARQ
 TKENVEVSFDDFSFGFIPKIQANMDPFRHRIAPRVVSGAFERVMMLCNVLKSK
 RSHVQHLMTKLLKHNLAGVDIGLDTGNQIGDTLVGKQYISFODLPQFTPEI
 FMKVSANKMQIKKHLKHEQVIEQATQYKTLHTNQILGAVGQLQFVEFHEMRKN
 EYNDVMPVPGVKRTARWDIENEDQITDKMNTSRSLVKDRYDDLVLFPENEFATRW
 EEKFEIKLYSL"
 complement(3009..3013)
 /gene="RF3"
 complement(3085..3091)
 /gene="RF3"
 complement(3109..3114)
 /gene="RF3"
 complement(3244..4784)
 /gene="mure"
 complement(3244..4722)
 /gene="mure"
 /codon_start=1
 /transl_table=11
 /product="UDP-N-acetylmuramyl-tripeptide synthetase"
 /protein_id="CAA74740.1"
 /db_xref="GI:3256223"
 /db_xref="SPTREMBL:O86491"
 /translation="MQVRLCKKVKRVKLVGLSEQOIDDITDTSRTAREGSIFVASGVY
 TVDSHKFCQNGCKLVKLVNKEQSLPANVQVVPDTRVASTLAHTLYDPSHQL
 VYFGVTGNTKSTIATMHLIQRKQNSAYLGTNGFOINETKTGANTPTPTVSLTK
 KTKEAVDAGAESSTLEVSCHGLVGLRSLVDFVALFSLTQDHLDFHCTMEAYGHAK
 SILFSLQEDLSKEXYVLNNDSDSEYLRTVTPVEYSYGLDEAQAQWAKNIQESLQ
 GVSDFVTFPGYPKSPVGFNININIAAMIAVMSKSLSETIILKAVENLEPVEGR
 LEVLDPSPILDIIDYAHADPMKNLIDAVQPVKOKLIFLVGMAGERDLTKTPMGR
 VACRADYVIFTPDNANDPDKMLTAEAKGATHQNYIEFDDRAEGIKHAIDIAEPGDT
 VYLASKGREPYQIMPGHKVPHRDDLIGLEAAKKYKFGGPPVD"
 complement(4736..4740)
 /gene="mure"
 RBS

-10_signal complement(4758..4763)
 /gene="mure"
 -35_signal complement(4779..4784)
 /gene="mure"
 gene 5086..6335
 /gene="ypfp"
 -35_signal 5086..5092
 /gene="ypfp"
 -10_signal 5107..5112
 /gene="ypfp"
 RBS 5145..5150
 /gene="ypfp"
 CDS 5160..6335
 /gene="ypfp"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA74741.1"
 /db_xref="GI:3256224"
 /db_xref="SPTREMBL:O86492"
 /translation="MYTONKILITGSGNGHMOVTQSIYNQINDMNDLHLSVIEHD
 LPEAHPLITSICKWYINSFKYFRNMYKGFYSRDPKLDKCFYKYGLNKLINLLIK
 EXPDLILLFPPTVMVLEQFNIPVATVMDIRLKNWILPPISTRIYVATKFTKQ
 DFIDVGDIDPSTVKVTGIPIDNAFETPINQKWLIDNNLDPDKOTILMSAGATGVSQGF
 MITDILAKSANAQVVMICGSKELKSLTAKFLKRLMYLILGTYTKHNEWMASQQL
 MITKPGGITIEGFARCIPIELNPAQGELENAFYFEKGFGLADTPEEAIKIVAS
 LTNGNEQLTNMISTWEODKIKYATOTICRDLDLIGHSSQPOEYIGKVPYARFVK"
 BASE COUNT 2517 a 1452 c 1130 g 2668 t 24 others
 ORIGIN

Query Match 98.9%; Score 964; DB 3; Length 7791;
 Best Local Similarity 99.9%; Pred. No. 1.2e-162;
 Matches 975; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atgggtactcaaaataaaagatattgattacttgcttcggtcattcggtcaagcgtcatatg 60
 DB 5160 ATGGTTACTCAAAATAAAAGATATTGATTACTGGCTCATTCGGTACGCTCATATG 5219
 QY 61 caagttacacagatcgtaactcaactaatatgatatgaatagaccatttaagcgtc 120
 DB 5220 CAAGTTACACAGAGTATCGTTAATCACTTAATGATGATGATGATGATGATGATG 5279
 QY 121 attgagcacgatttatttggagctcatcccaattttgactctctattgtgaaaaatgg 180
 DB 5280 ATTGAGCACGATTATTATTATGGAAGCTATCCCAATTTTGACTTCTATTGTAATAAATGG 5339
 QY 181 tatatcaatagcttttaaaatatttagaataatgacaaaggggttttattacagccgcca 240
 DB 5340 TATATCAATAGCTTTAAATATTATTTAGAAATATATGACAAAGGGTTTTATTACAGCCGCCA 5399
 QY 241 gataaactagacaaatgtttttacaaatactatggacttaataagtaattattattg 300
 DB 5400 GATAAAGTACACAAATGTTTTACAAATACTATGACTTAATAAGTAAATTAATTATTATG 5459
 QY 301 ataaagaaagccagatttaattattattacagtttctacacaggttatgtcggtacta 360
 DB 5460 ATAAAGAAAGCCAGATTAAATATTATTAACTTCTTACACACAGTATTATGTCGGTACTA 5519
 QY 361 actgagcaatttaacattattccagttgctacagtgatgacagactatgcttaccat 420
 DB 5520 ACTGAGCAATTTAACATTAAATATTATCCAGTGTGCTACAGTATGATGATGATGATGATG 5579
 QY 421 aaaaactggattacgcgctattcaacaagatatattgtggcaacaaagaacgaacaa 480
 DB 5580 AAAAAGTATTACGCGGTTATTCAACAAGATATTATGTCGCAACAAAGAAAGAAACAA 5639
 QY 481 gacttcagacgttaggtattgatctctcaacagtttaaaagtgacaggtattctcttattg 540
 DB 5640 GACTTCATGACGTAGGTATTGATCTCTCAACAGTTAAAGTGAAGTATTCCTATTGAT 5699
 QY 541 aacaaatttgaaacgcttatttaataaagcagtggtttaaataagacaaacacttagatcca 600
 DB 5700 AACAAATTTGAAACGCTTATTATTAATAAAGCAGTGGTTAATAGACAACTTAGATCCA 5759

LOCUS AF270394 3493 bp DNA BCT 01-AUG-2000
DEFINITION Staphylococcus epidermidis strain SR1 clone step.4046a02 genomic
sequence.
ACCESSION AF270394
VERSION AF270394.1 GI:9624308
KEYWORDS
SOURCE
ORGANISM Staphylococcus epidermidis.
Staphylococcus epidermidis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE 1 (bases 1 to 3493)
AUTHORS Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M.,
Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
Listenbee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
Furdon, P.J.
TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3493)
AUTHORS Taylor, J., David, Kimmerly, W.J., Nelsen, A.J., Godlevski, M.M.,
Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
Listenbee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
Furdon, P.J.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
source
Location/Qualifiers
1..3493
/organism="Staphylococcus epidermidis"
/strain="SR1"
/db_xref="taxon:1282"
/clone="step.4046a02"
BASE COUNT 1172 a 534 c 547 g 1240 t
ORIGIN

Query Match 65.3%; Score 636.2; DB 2; Length 3493;
Best Local Similarity 78.9%; Pred. No. 3.2e-104;
Matches 758; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 1 atggttactcaaaataaaagatattgattactggtcctcattcgttaacggtcattg 60
DB 384 ATGGTTACTCAAAATAAAAGATATTGATTACTGGATCATTTGGTAACGGTCATATG 443

QY 61 caagttacacagatgattcctcaactaatgatatgaatcctagaccatttaagcgctc 120
DB 444 CAAGTCACCAAGTATTGTCACCAATTTGAATGAGATGAATCTCAATCATTTATCAGTC 503

QY 121 attgagcagcatttatttgaagctcctcaattttgactctcatttcttataaaatgg 180
DB 504 ATTCAACATGATTGTTTATGGAAGCTCATCCAAATTTATGACTTCATATGTAAGAAATGG 563

QY 181 tatataaatgatttataatattttgagaataatgataaggggttttattacagcgccca 240
DB 564 TATATCAATAGCTTTAAATATTTTGAATAATACATATAAACGATTATTACTATAGTCGCCCT 623

QY 241 gataaactagacaattgttttaccataactatgacttaataagtttaattattatg 300
DB 624 AATGAGTCGATAAATGTTTTTATAAATATTGATTAATAAATCACTCATCACTTACTT 683

QY 301 ataaagaaagccagatttaattattataacggtttcctacaccagttatgctcggtacta 360
DB 684 ATTAAGAAAGCCGATGCTCATATATTAAACATTTCCAACACCTGTGTATGTCAGTGTG 743

QY 361 actgagcaatttaacatttaattccagttgctacagtgatgacagactatgcttacctat 420
DB 744 ACCGACAAATTTAAATATAATATCCCTATTGGCAGAGTTATGACAGATTATCGCATGCA 803

QY 421 aaaaactggattacgcgctattcaacaagatatattgtggcaacaaagaaacgaaacaa 480
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 804 AAAAATGGATTACACCATATTACAAAGATATTATGTAGCAACAAAGATACTAAGAT 863
QY 481 gacttcatacagcgtaggtattgatccttcacaggttaaaagtgcaggtattcctcattgat 540
DB 864 GATTTTCATTGAAGCTGGTGTCTCTGCTCATATATTAAAGTGCAGGGCATTCCTATTGCT 923

QY 541 acaaaatttgaaacgcctatttaatacaaaagcagctggttaataagacaacacttagatcca 600
DB 924 GATAAATTTGAGAAATCTATTGATTAAGAAGAAATGTTATTCGCAACAACATTTAGACCCT 983

QY 601 gataagcaaaactatttttaagtgcagctggtgcattttgggtgtatcctaaaggttttgacacg 660
DB 984 TCAAAACCTTACTATATTAAATGTCAGCAGCTGTCATTTGGTGTTCAAAGGCTTTGACTAT 1043

QY 661 atgattactgatattatagcgaagagtcgaaatgcacaagtagttatgatttgggtgaag 720
DB 1044 ATGATTAATAATATTTTAGAAAAAGTCCAAATTCCAAGTGTGTCATGATTTGTGGACGT 1103

QY 721 agcaaaagcgttaaagcgttctttaaagcagctaaagttttaaataaacgagaatgtattgatt 780
DB 1104 AGTAAGCAACTTAAGCTTCATTAAGAGCTAAGTTCAAGATTAATCCAAAGTGTAAATA 1163

QY 781 ctaggttataccaaacacacatgaatgaatggtgcatcgaatcgaacttatgatattacgaaa 840
DB 1164 TTAGGATATACAAATCACATGAATGATGGATGGCATCAAGCAACTAATGATTACAAA 1223

QY 841 cctggtggtatcacataaactaaaggttcgcgcgttggttatcccaatgattttcctaaat 900
DB 1224 CCTGGTGGTATCACAAATTTCCGAAGGACTTAGTCGTTGTTATTCCTATGATTTTAAAC 1283

QY 901 cctgcaacctggtcagagcttgaaaaatgccttttaccttgagaagaaaggtttgtgtaaa 960
DB 1284 CCTGCACCCGGTCAAGAACTTGAAATGTCATATTACTTTGAAAGTAAGGATTTGGAAA 1343

QY 961 a 961
DB 1344 A 1344

RESULT 5
AR106079/C
LOCUS AR105079 2600 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6103507.
ACCESSION AR105079
VERSION AR105079.1 GI:12820144
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2600)
AUTHORS Shilling, L. Kathleen, Wallis, N. Gail and Zabierowski, S.
TITLE Mure
JOURNAL Mure
FEATURES
Patent: US 6103507-A 1 15-AUG-2000;
Location/Qualifiers
source
1..2600
/organism="unknown"
BASE COUNT 393 a 369 c 494 g 844 t
ORIGIN

Query Match 44.8%; Score 437; DB 9; Length 2600;
Best Local Similarity 100.0%; Pred. No. 1e-68;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttactcaaaataaaagatattgattactggtcctcattcgttaacggtcattg 60
DB 557 ATGGTTACTCAAAATAAAAGATATTGATTACTGGCTCAATTCGTAACGGTCATATG 498

QY 61 caagttacacagatgattcgttaacttaactaatgatgatgaatcagaccatttaagcgctc 120
DB 497 CAAGTTACACAGATATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 438

QY 121 attgagcagcattttatttggaagcctatccaatttgcactctatttggtaaaaaatgg 180

```
Db 437 ATTGAGCAGCATTTATTATGAGAGCTCATCCCAATTTGACTTCTATTGTAATAAATGG 378
QY 181 tataatagctttaaatatttagaataatgtacaaaaggggtttattacagcgccca 240
Db 377 TATATCAATAGCTTAAATATTTAGAAATATGTACAAAGGGTTTATTACAGCGGCCA 318
QY 241 gataaactagacaaatgtttttacaatactatgacttaataagtttaataattattg 300
Db 317 GATAAATAGACAATGTTTTTACAATACTATGAGACTTAATAAGTTAATTAATTATTG 258
QY 301 ataaagaagaagccagatttaattattacgttttctacacagttatgctggtacta 360
Db 257 ATAAAGAAAAGCCAGATTTAATATTATTAAACGTTTCTTACACAGTATGCGGTACTA 198
QY 361 actgagcaatttaacattatccagttgctacagtgatgacagactatcgcttacat 420
Db 197 ACTGAGCAATTTACATAATATATCCAGTTGCTACAGTGATGACAGACTATGCGTTACAT 138
QY 421 aaaaactggattacgcc 437
Db 137 AAAAAGTGGATTACGCC 121

RESULT 6
E35634/c 2600 bp DNA PAT 07-FEB-2001
LOCUS E35634
DEFINITION Mure.
ACCESSION E35634
VERSION E35634.1 GI:13019110
KEYWORDS JP 1999225779-A/1.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2600)
AUTHORS Nicola,G.W.L.L. and Schilling,S.Z.
TITLE Mure
JOURNAL
COMMENT Patent: JP 1999225779-A 1 24-AUG-1999;
SMITHKLINE BEECHAM CORP, SMITHKLINE BEECHAM CORP PUBLIC LTD CO
OS Unidentified
PN JP 1999225779-A/1
PD 24-AUG-1999
PR 25-SEP-1998 JP 1998-09356
PF 25-SEP-1997 US 60/059956,10-DEC-1997 US 08/988251 PI
NICOALA G WALLIS, LISA K SCHILLING, SUSAN ZABIAROSUKI PC
C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00, PC
A61K31/00,
PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K35/74,
PC A61K38/00,
PC A61K39/085, A61K39/395, A61K39/395, A61K45/00, A61K48/00,
C07K16/12, PC C12N9/00,
PC C12Q1/68, G01N33/53/C12P21/08, (C12N15/09, C12R1:46), C12N15/00,
PC A61K37/02,
PC (C12N15/00, C12R1:46)
CC Strandedness: Double;
CC Topology: Linear;
FH Key
FT source
FT Location/Qualifiers
FT /organism="Unidentified".
FT source
FT Location/Qualifiers
FT 1..2600
FT /organism="unidentified"
FT /db_xref="taxon:32644"
BASE COUNT 893 a 369 c 494 g 844 t
ORIGIN

Query Match 44.8%; Score 437; DB 10; Length 2600;
Best Local Similarity 100.0%; Pred. No. 1e-68;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttactcaaaataaaagatattgattattactggtcattcggtacaggtcatatg 60
|||||
```

```
Db 557 ATGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498
QY 61 caagttacacagagtcgttaactcaacttaataatgatatgaatctagaccatttaagcgtc 120
Db 497 CAAGTTACACAGAGATGCGTTAATCAACTTAATGATGATGAATCTAGACCATTAAAGCGTC 438
QY 121 attgagcacgatttatttattggaagctcatccaaattttgacttctatttgaataaaatg 180
Db 437 ATTGAGCAGCATTTATTATGGAAGCTCATCCCAATTTGACTTCTATTGTAATAAATGG 378
QY 181 tatataatagctttaaatatttttagaataatgtacaaaaggggtttattacagcgccca 240
Db 377 TATATCAATAGCTTAAATATTTAGAAATATGTACAAAGGGTTTATTACAGCGGCCA 318
QY 241 gataaactagacaaatgtttttacaatactatgacttaataagtttaataattattg 300
Db 317 GATAAATAGACAATGTTTTTACAATACTATGAGACTTAATAAGTTAATTAATTATTG 258
QY 301 ataaagaagaagccagatttaattatttatttaacgttttctacacagttatgctggtacta 360
Db 257 ATAAAGAAAAGCCAGATTTAATATTATTAAACGTTTCTTACACAGTATGCGGTACTA 198
QY 361 actgagcaatttaacattatccagttgctacagtgatgacagactatcgcttacat 420
Db 197 ACTGAGCAATTTACATAATATATCCAGTTGCTACAGTGATGACAGACTATGCGTTACAT 138
QY 421 aaaaactggattacgcc 437
Db 137 AAAAAGTGGATTACGCC 121

RESULT 7
AX016296 1149 bp DNA PAT 07-SEP-2000
LOCUS AX016296
DEFINITION Sequence 1 from Patent WO9949052.
ACCESSION AX016296
VERSION AX016296.1 GI:10041859
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE 1 (bases 1 to 1149)
AUTHORS Zaehringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.
TITLE Processive glycosyltransferase
JOURNAL Patent: WO 9949052-A 1 30-SEP-1999;
FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
(DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
UND VERWER (DE)
FEATURES
source
source 1..1149
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 380 a 217 c 244 g 308 t
ORIGIN

Query Match 12.1%; Score 118; DB 9; Length 1149;
Best Local Similarity 48.5%; Pred. No. 7.6e-12;
Matches 458; Conservative 0; Mismatches 465; Indels 21; Gaps 4;

QY 13 aataaaaaagattgattattactggtcattcgttaacggttcattgcaagttacacag 72
|||||
Db 10 AATAAAGAGTATTAAATTTTGACTGCAAAATACGGAATGACATGTGCAGGTAGCCAAA 69
QY 73 agtatcggttaactcaacttaataatgatatgataccatttaagcgtcattgagacagat 132
Db 70 ACATTTTATGAACAATGTGTACGCGCTCGGCTTTTCAGCATGTAAACAGTTTCT- 123
|||||
QY 133 ttatttatgaaagctcatcaattttgacttcttattgttttaaaaaatggtatatacagc 192
|||||
Db 124 TTGTACCAGAGTCAAAATCCGATTGTTTCAGAGTAACTCAATACCTTTATTAAAAAGC 183
|||||
```

QY	193	tttaaatatttagaataatgtacaaaggggtttttatt-----acagcgccccagat	243	ORGANISM	Bacillus subtilis
Db	184	TTTCTAATCGGGAACAGATTTTATCGTTGTTTATACGGAGTTGACAAATCTATAAT	243	REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;
QY	244	aaactagacaaatgtttttacaaataactactgacttaataagtttaattattatgata	303	AUTHORS	Bacillus/Staphylococcus group; Bacillus.
Db	244	AAACGTAATAATTAATTTACTTTAAATGGGTAATAAAGATTTGGCGCAACTTGTCGAT	303	TITLE	1 (bases 1 to 35040)
QY	304	aaagaaagccagatttaataattattacagtttccctacaccagttatgtcggtactaaat	363	JOURNAL	Iwakura, M., Kawata, M., Tsuda, K. and Tanaka, T.
Db	304	GAACATCAGCCGCAATTAATTAATATACATTTCCGATGATCGTCGCGCAATACAGA	363	MEDLINE	Nucleotide sequence of the thymidylate synthase B and dihydrofolate
QY	364	gagcaatttaacatttaattccagttgtctacagtgatgacagactatcgcttacataaa	423	REFERENCE	reductase genes contained in one Bacillus subtilis operon
Db	364	CGCGAACTGGAAGATCTCTACCTTCAAGTTATGACTGATTTTGTCTTCATAAA	423	AUTHORS	Gene 64 (1), 9-20 (1988)
QY	424	aactgattacgcgcgtattcaacaagatatatttggtggcaaaaagaaacgaaacagac	483	TITLE	88284366
Db	424	ATTGTGGTTTCACGAAACGTCGTAATAATATTGTGGCGCACAGATTACGTGAAGGAAAA	483	JOURNAL	2 (bases 1 to 35040)
QY	484	ttcatagacgtaggtattgatccctcaacagttaaagtgacaggtattcctatttgataac	543	REFERENCE	Sorokin, A.V., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S.D.
Db	484	CTGCTGGAGATCGCACTCATCCAAACATGTAAANAATCACAGGAATTCATCAGCCGC	543	AUTHORS	and Serror, P.
QY	544	aaattgaaacgcctatttaatacaaaagcagtggttaatatagacaaacacttagatccagat	603	TITLE	Sequence analysis of the Bacillus subtilis chromosome region
Db	544	CAATTTGAAGAATCCATGCTGTGGCCGATATATAAAGTACAAATCTTTCACCAAC	603	JOURNAL	between the sera and kdg loci cloned in a yeast artificial
QY	604	aagcaactatttaagtgcagtgctgattgggtggtatcctaagattttgacacgatg	663	FEATURES	chromosome
Db	604	AAAAAGTCTCTGATCATGCGGAGTGTCTACGGTGTATTAAAGAACGTAAGAGCTG	663	source	Microbiology (1996) In press
QY	664	attactgatatttagcgaagtcgaatgcacaagtgattgatgttggtgtaagagc	723	gene	1. 35040
Db	664	TGC---GAAACCTTCTCAAGGATGACCAAGTGCAGTAGTTGCTGTGCGGGAATAAT	720	CDS	/organism="Bacillus subtilis"
QY	724	aaagactaaagcgttcotttaacagtgattgatttaataattacagagaattgattga---tt	780	terminator	/sub_strain="168"
Db	721	ACGGCTTTAAAGAAATCTTTAGTGGCTTGAAGCGGAATAATGGTGACAAATTTAAAGTT	780	gene	/strain="Marburg"
QY	781	ctaggttataccaaacacatgaatgaatggtgcatcgaactaatgattacgaaa	840	CDS	/db_xref="taxon:1423"
Db	781	CTGGGTATGTGGGCGCAATGTAGAGCTATTTCCGATCACAGATTGCATGATTACCAAG	840	terminator	/clone_lib="P. Serror"
QY	841	cctggtgtatcacaaataactgaaggttttcgcgcgtttgattccaatgattttcctaaat	900	gene	/map="205 degrees"
Db	841	CCGCGCGCATTAATTGACAGAAGCCACAGCCATTGGAGTGCCTGCTCATCTCTGTACAAA	900	CDS	1. 1095
QY	901	cctgcacctcgtcaagcgttgaataatgccttttactttgaaga	944	terminator	/gene="ypwA"
Db	901	CCCGTCCTGGCCGGAAGAAAGAAATGCAAACTTCTTTGAAGA	944	gene	<1. 1095
RESULT	8				/gene="ypwA"
BACVACA/c					/note="putative"
LOCUS					/citation=[3]
DEFINITION	BACVACA 35040 bp DNA BCT 10-APR-1996				/gene="xpt"
ACCESSION	L77246				/gene="xpt"
VERSION	L77246.1				/db_xref="GI:1256617"
KEYWORDS	DNA-binding protein; adenine phosphoribosyltransferase; bcsA gene; bsaA gene; cold shock protein; cspp gene; degr gene; dfrA gene; dihydrofolate reductase; dihydroxyacid dehydratase; exodeoxyribonuclease; homoserine O-succinyltransferase; ilvA gene; ilvD gene; metB gene; naringenin-chalcone synthase; stress-related protein; threonine dehydratase; thyB gene; thymidylate synthase; transport protein; xpt gene; ypaQ gene; ypbQ gene; ypbR gene; ypbS gene; ypcp gene; ypdP gene; ypdQ gene; ypeP gene; ypeQ gene; ypiP gene; ypgQ gene; ypgR gene; yphP gene; ypiP gene; yjpP gene; yjPQ gene; ypmR gene; ypmS gene; ypmT gene; ypnP gene; ypnQ gene; yppP gene; yppQ gene; yppQ gene; yppq gene; ypwA gene.				/ec_number="2.4.2.7"
SOURCE	Bacillus subtilis.				/function="purine biosynthesis"

```

CDS
2007..3323
/feature="pyrimidine salvage"
/notes="26% of identity to the Bacillus caldolyticus uracil
transport prot. . . : putative"
/citation=[3]
/codon_start=1
/transl_table=11
/product="transport protein"
/protein_id="AAA96612.1"
/db_xref="GI:1256618"
/transl_table=11
/translation="MRNGFKTSLGQIHLVLAAGAIIVPLIVKAMGLTVEQLTYL
VSDIFMCGVATLIVNRSRFFGIGVVLGCTFTAVSPMIAIGSYGVSTVYGSIIA
SGILVLSIFFKLVSGFFPVVGTGVTIIGITLPMVNMNMGAGEGADFGDLSNL
ALAPTIVLSIVLYRTTIGFIKSVSILIGLITFTAYFMGKVQFONVSDAAVVOVMIQ
PFYEGAPSEHAAPILDAIVAVISVESTGVFALGDLNRRLEIDLSKGRAGELA
VLGGAPNAPYAFQSONGLVQLTGIKKNNAVIVGVILMAGLPFKTAFTTIIPS
AVLGGAMVAFYVGIKMLSRIDFAQENLLIVACSVGLGLGVTVVVPDIFKQLPS
ALTLTNGIVAGSFATVNLNIVYVFSKAKKIEQADLAEQKTA"
3322..3341
/feature="putative"
/citation=[3]
3397..4494
/feature="bcsA"
3397..4494
/feature="bcsA"
/EC_number="2.3.1.74"
/notes="25.6% of identity to the Lycopersicon esculentum
(tomato) naringenin-chalcone synthase; alternate gene name
ypbP; putative"
/citation=[3]
/codon_start=1
/transl_table=11
/product="naringenin-chalcone synthase"
/protein_id="AAA96613.1"
/db_xref="GI:1256619"
/transl_table=11
/translation="MAFILSIGSLPAYNVNQKAEFFRYMFQHFQKIDIRLLSFK
NQIHSQFVKPIEYWKESHSFEKQIYIEETLKHRSRAVRECLSHPHFQBAIYE
KBAVFFSVSTPSIEARLNLNLPSPYTRIPITWIGLCAGAGSLARRAAEYCKA
YPFAVLVIAELCSLTFOPEDKTKNLIGTSLFGDIAAALLCGEKADRVRSKLIFA
PKIMDAOVLMLKOSDVMGDFDQGFVIFSRDIPTLVEKWLKTNVQIFLDKHLKLSF
HDSVFLAHFGGKVIDAYIKSLGLSEKLSAQSLQIKHGNSSATILYIVIKHLQNL
GHKAEKGLIGLPGFSSELLFSWEKA"
4498..5004
/feature="ypbQ"
4498..5004
/feature="ypbQ"
/notes="33.3% of identity in 60 aa with the Saccharomyces
cerevisiae farnesyl cysteine carboxyl-methyltransferase;
putative"
/citation=[3]
/codon_start=1
/transl_table=11
/protein_id="AAA96614.1"
/db_xref="GI:1256620"
/transl_table=11
/translation="MFLLIAILIVQRAEMAVARQNEQVKKQGAIEFGESHYPYII
TMHILFLSLIAELNKPSSWNLGAAVILSVQIRYVWALCSGAYWNKILVVP
GVELVKGKPYKMKHPNAVYVILEILLIPLLYQAYVTMCLSFNFNAVLLSVIRAEQK
ALQBSYVK"
5340..8921
/feature="ypbR"
5340..8921
/feature="ypbR"
/notes="26.7% of identity in 165 aa to a Thermophilic
bacterium hypothetical protein 6; putative"
/citation=[3]
/codon_start=1
/transl_table=11
/protein_id="AAA96615.1"
/db_xref="GI:1256621"
/transl_table=11
/translation="MTQNRKELLHKTGELYKOFIENQDQRAAKLAAMVKKAADEV
YIAPTGHYSAGKSLNCLMLNIPSPISPTSANLVIRNKEKRVRLHPTDGCACEL
EGYQKDKVQYCKDGEIESVEIFDRYTEIDSGVVIDPFGIDSTDDAHLFSAAIL
terminator
gene
CDS
3397..4494
/feature="bcsA"
3397..4494
/feature="bcsA"
/EC_number="2.3.1.74"
/notes="25.6% of identity to the Lycopersicon esculentum
(tomato) naringenin-chalcone synthase; alternate gene name
ypbP; putative"
/citation=[3]
/codon_start=1
/transl_table=11
/product="naringenin-chalcone synthase"
/protein_id="AAA96613.1"
/db_xref="GI:1256619"
/transl_table=11
/translation="MAFILSIGSLPAYNVNQKAEFFRYMFQHFQKIDIRLLSFK
NQIHSQFVKPIEYWKESHSFEKQIYIEETLKHRSRAVRECLSHPHFQBAIYE
KBAVFFSVSTPSIEARLNLNLPSPYTRIPITWIGLCAGAGSLARRAAEYCKA
YPFAVLVIAELCSLTFOPEDKTKNLIGTSLFGDIAAALLCGEKADRVRSKLIFA
PKIMDAOVLMLKOSDVMGDFDQGFVIFSRDIPTLVEKWLKTNVQIFLDKHLKLSF
HDSVFLAHFGGKVIDAYIKSLGLSEKLSAQSLQIKHGNSSATILYIVIKHLQNL
GHKAEKGLIGLPGFSSELLFSWEKA"
4498..5004
/feature="ypbQ"
4498..5004
/feature="ypbQ"
/notes="33.3% of identity in 60 aa with the Saccharomyces
cerevisiae farnesyl cysteine carboxyl-methyltransferase;
putative"
/citation=[3]
/codon_start=1
/transl_table=11
/protein_id="AAA96614.1"
/db_xref="GI:1256620"
/transl_table=11
/translation="MFLLIAILIVQRAEMAVARQNEQVKKQGAIEFGESHYPYII
TMHILFLSLIAELNKPSSWNLGAAVILSVQIRYVWALCSGAYWNKILVVP
GVELVKGKPYKMKHPNAVYVILEILLIPLLYQAYVTMCLSFNFNAVLLSVIRAEQK
ALQBSYVK"
5340..8921
/feature="ypbR"
5340..8921
/feature="ypbR"
/notes="26.7% of identity in 165 aa to a Thermophilic
bacterium hypothetical protein 6; putative"
/citation=[3]
/codon_start=1
/transl_table=11
/protein_id="AAA96615.1"
/db_xref="GI:1256621"
/transl_table=11
/translation="MTQNRKELLHKTGELYKOFIENQDQRAAKLAAMVKKAADEV
YIAPTGHYSAGKSLNCLMLNIPSPISPTSANLVIRNKEKRVRLHPTDGCACEL
EGYQKDKVQYCKDGEIESVEIFDRYTEIDSGVVIDPFGIDSTDDAHLFSAAIL

```


QY 484 ttcatagacgttaggtattgatctctcaacagtttaaaagtcagagtgattcttattgataac 543
 Db 14453 CTGCTGAGATCGCATCTATCCAAAGCAATGTAAATATCAGGAATTCATCAGGCCG 14394
 QY 544 aaatttgaacgctatttaatacaaaagcagtggttaataagacaaacttagatccagat 603
 Db 14393 CAATTTGAAGAAATCCATGCTGTGGCCGATATATAAAAGTACAATCTTTACCAAAAC 14334
 QY 604 aagcaactattttaatgtcagctggtgcatttggtgtatcttaaaagttttagacagtg 663
 Db 14333 AAAAAGTCTTCTGATCATGCGAGGTCACCGTGATTAAGAACGTAAGAAGCTG 14274
 QY 664 attactgatattatagcgaaagtgcacaaatgcacaaatgattgattggtgtaagagc 723
 Db 14273 TGC---GAAACCTGTCAAGGATGACCAAGTGCAGTAGTGTGCTGCGGGAAAAAT 14217
 QY 724 aagagctaaagggtttttaacagctaaagttaaaatcaacgagaatgtatttga---tt 780
 Db 14216 ACGGCTTTAAAGAAATCTTTGAGTGCCTTGAAGCGGAAATGGTGACAAATTAAGAGTT 14157
 QY 781 ctatgtttaccacaaacatgaatgaatgagtcacatgaatgaatgaatgaatgaatgaat 840
 Db 14156 CTGGCTATGTGAGCCCATGATGACATTTTCGGATCATCAGATTCATGATACCAAG 14097
 QY 841 cctggtgtatcacaaataaactgaagtttccgcccgttgatttccaaatgatttccctaaat 900
 Db 14096 CCGGGCGCATTTACTTTGACAGAAAGCACAGCCATTCGAGTGCCTGTCTGTACAAA 14037
 QY 901 cctgcacctgtcaagagcttgaatgaaatgcttttaacttttaaga 944
 Db 14036 CCGTGCCTGGCCAGGAAAGAAATGCAAACTCTTTGAAGA 13993

RESULT 9

BSUB0012 213680 bp DNA BCT 26-NOV-1997
 LOCUS Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220.
 ACCESSION 299115 AL009126
 VERSION 299115.1 GI:2634478
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus subtilis.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
 Bacillus/Staphylococcus group; Bacillus.
 1 (bases 1 to 213680)
 Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruchi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Chim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, S.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F., Itaya, M., Jones, L., Joris, B., Karanata, D., Kasahara, Y., Klaier-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardin, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maule, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portet, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Serron, S.J., Serron, P., Shin, B.S., Soldo, B., Sorokin, A., Taccori, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Takakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,

Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.
 The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*
 Nature 390 (6657), 249-256 (1997)
 98044033
 2 (bases 1 to 213680)
 Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
 Direct Submission
 Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48
 Location/Qualifiers
 1. 213680
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 complement(90..782)
 /gene="yoqL"
 complement(90..782)
 /gene="yoqL"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13978.1"
 /db_xref="GI:2634479"
 /db_xref="SPTREMBL:O34581"
 translation="MDSKKKIINGYHKNICISCKSLPATEENFYSVKKNKDLHSY CKAVLKAKESMLKNDYQQLERMRNLLPGMKAKKYNLSLKKKTQQIWEKKNK LKLNRYLQORDAHKKHNTDVOWCKDYFNKCSYGLKIEDHKILFKGTYSIDFH KEVDHKGANDISNCIPACKSCNSKHDFAFEWYNSSNKNFTSSERLLAIKEWLNRK EERQSEWRTG"
 complement(822..1025)
 /gene="yoqK"
 complement(822..1025)
 /gene="yoqK"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13978.1"
 /db_xref="GI:2634480"
 /db_xref="SPTREMBL:O34326"
 translation="MKLIERTYEERLYISDDIEARKHFVSELKPKGWNVDCWVGI DLNLIKQFYRFKRELTDVFNFR"
 complement(1045..1560)
 /gene="yoqJ"
 complement(1045..1560)
 /gene="yoqJ"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13979.1"
 /db_xref="GI:2634481"
 /db_xref="SPTREMBL:O34359"
 translation="MKNPTMLKDKDLAVIEELITKENKRYFITGAGLTQQAACWC VHLKKHPHINKIATPEKODKYWSADOKMYKPMVNADEILINVELDKYKVSFGE KGESEPAKMQKNEYMDHSEIAVAVDYSKSGTRNCLNIAKTYLGHQLWRLHPDF NFELDITYFVG"
 complement(1771..1965)
 /gene="yoqI"
 complement(1771..1965)
 /gene="yoqI"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB13980.1"
 /db_xref="GI:2634482"
 /db_xref="SPTREMBL:O34339"
 translation="MKSIILPEENTKITVGRPINDESNKVIAYIDYREEPEAFW"

```
gene
CDS
    HLSGNDLFVDNHEVIVEVE"
    complement(2014..2466)
    /gene="yoqH"
    CDS
    complement(2014..2466)
    /gene="yoqH"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13981.1"
    /db_xref="GI:2634483"
    /db_xref="SPTREMBL:O34999"
    /translation="MKRFILVLSFLSIIVAYPIOTNASPMPCSVILEPVDKLNKNAG
    VALIYKQLAPPSAARTNISILAVHLPPSPFCNVDYSYEGFATKPGELISWRFLYTP
    EESPWAGRIDIISEMKNVKQVRLSNSSTOKLGPSILTKNIESCY"
    complement(2549..2806)
    /gene="yoqG"
    CDS
    complement(2549..2806)
    /gene="yoqG"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13982.1"
    /db_xref="GI:2634484"
    /db_xref="SPTREMBL:O35030"
    /translation="MEKNLLKEITIFDLNKKIRPKVQVVTWKGTMEYTHNGEVII
    NGEKPYNVVDKEGYVGHCHVNALDLKNYPDSLIVEIKSK"
    complement(2851..3054)
    /gene="yoqF"
    CDS
    complement(2851..3054)
    /gene="yoqF"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13983.1"
    /db_xref="GI:2634485"
    /db_xref="SPTREMBL:O34584"
    /translation="MKFAPMDKVKFKTASHLNKLRLTLKRVPELDDPLLGEWEFEED
    GLRFDEENVEFVARPKHENWD"
    complement(3063..3227)
    /gene="yoqE"
    CDS
    complement(3063..3227)
    /gene="yoqE"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13984.1"
    /db_xref="GI:2634486"
    /db_xref="SPTREMBL:O34429"
    /translation="MENNLYGVYSTRSKRKFGEIPEPSKTARKELFNRIQDAYKW
    RFEIRKIKRK"
    complement(3281..4036)
    /gene="yoqD"
    CDS
    complement(3281..4036)
    /gene="yoqD"
    /function="unknown"
    /note="similar to phage-related DNA-binding protein
    anti-repressor"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13985.1"
    /db_xref="GI:2634487"
    /db_xref="SPTREMBL:O34449"
    /translation="MESLYIVIEQNGQLLDSREVAEMVGRKHTDLLRSIDGVAILL
    NAKLSRVEFFLESTYKDATGSLKHFHLTRKGDWANKMTGAKVLFITAQYVSKFEE
    MEKALRPSLIDYLDNMEDERAIAVTERKAKRELQEOLTLAEKPKVEKVDRLNFD
    GLMKIGVAKAIGIKMGQNNLFRLENKVLIDGTNKNAPYQKYVERGFQVKQET
    SVGIKTLVTPKGADFIYDLKKHGHKREIAS"
    complement(4077..4484)
    /gene="yoqC"
    CDS
    complement(4077..4484)
    /gene="yoqC"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13986.1"
    /db_xref="GI:2634488"
    /db_xref="SPTREMBL:O34834"
    /translation="MEVNVKTNQREKFIIRNGIPYDELDTQMHLIDILNFKIGLKRH
    CCFGHKPYEEIOVMEFEDVNIKEIDOILELAELAGREWKQLQSFSKWARESPLMFNWS
    LVLSKFRNPEDPNKYRLRSVEEFESYAAKK"
    complement(4491..4829)
    /gene="yoqB"
    CDS
    complement(4491..4829)
    /gene="yoqB"
    /function="unknown"
    /codon_start=1
    /transl_table=11
    /protein_id="CAB13987.1"
    /db_xref="GI:2634489"
    Query Match 12.1%; Score 118; DB 2; Length 213680;
    Best Local Similarity 48.5%; Pred. No. 4.6e-12;
    Matches 458; Conservative 0; Mismatches 465; Indels 21; Gaps 4;
    QY 13 aataaaagattgattattactagtcctattcggttaacggtcatatgcaagtacacag 72
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110225 AATAAAAGAGTATTAATTTTACTGCTCAAAATTACGGAATGGACATGTGCAGGTAGCCAAA 110284
    QY 73 agtagcttaatacaacttaataatgaatcctagaccatttaagcgctcattgagcacag 132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110285 ACCTTTATGACAAATGTGTACGGCTCGGCTTCAGCATGTAAACAGTTTCT-----AAT 110338
    QY 133 ttatttgaagctcctccaaattttgactctctatttgaataaaatggtatacaatagc 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110339 TTGTACCAAGAGTCAAAATCGGATTTGTTTCAGAGGTAACCTCAATACCTTTATTTAAAAAGC 110398
    QY 193 tttaaatattttagaataatgtacaaagggtttatt-----acagcgccacagat 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110399 TTCTCAATCGGGAACAGTTTATFCGTTGTTTATTACGGAGTTTGACAAATCTATAAT 110458
    QY 244 aaactagacaaaatgtttttcaaaatactatggacttaataagtttaattattattgata 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110459 AAAGTAAATTCATAATTTACTTTTAAATGGTAATAAAAGATTGGCGCAACTTGTTCGAT 110518
    QY 304 aaagaaagccagatttaataatttaacgcttctcaccacacagttatgctggtactaac 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110519 GAACATCAGCCCGATATTTATTAATACATTTCCGATGATGTCGTCGCGGGAATACAGA 110578
    QY 364 gagcaatttaacataatctccagttgctacagtgatgacagactatcgcttacataaa 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110579 CGCGAAGTGGAAAGTCATTCCTACCTTCACGTTATGACTGATTTTGTCTTCATAAA 110638
    QY 424 aactgattacgcgcgtattccaacagatattatgtgcaacaaagaaagaaacaaagac 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110639 ATTTGGGTTACGAAAAACGTGGATAAATATTTATGTGCGACAGATTACGTGAAGGAAAAA 110698
    QY 484 ttcatacagtagtattgatctctcaacagtttaagtcagcagattcttctattgataac 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110699 CTGCTGGAGTCGGCACTCATCCCAAGCAATGTAATAATACAGGAATTCACAGCCCG 110758
    QY 544 aaatttgaacgcctatttaatacaaaagcagtggttaatagacaacacttagatccagat 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110759 CAATTTGAGAAATCCATGCTGTTGGCCGATATATAAAAGTACAATCTTTCACCAAAC 110818
    QY 604 aagcaactatttaagtctgctggtgatttgggttatcaaaagttttgacacacagtg 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110819 AAAAAGTGCCTTCATCATGGCAGTGTCTCACGCTGTATTAAAGAACGTAAGAGACTG 110878
    QY 664 attactaatattagcgaagaagtcgaatgcacaaatagttattggttggttaagagc 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110879 TGC---GAAACCTTGTCAGGATGACCAAGTCAAGATTTGTCGTGCGGGAATAAT 110935
    QY 724 aaagagctaaagcgttcttttaacagctaaagtttaataataacagagaatgtattga---tt 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 110936 ACGGCTTTAAAGAAATCTTTTGAGTCGCTTGAAGCGGGAATAATGTTGACAAATTTAAAGTT 110995
```


Db 1081 GAGATTCTGTTAAGGTTTCGAGGATTTCGAGACTCAATGAGAAATGATGGAGCTTG 1140
QY 822 tcaactatgattacgaacacctgggtggtatcacataaactgaagtttcgcccggttat 881
Db 1141 TGACTGCATCATCACAAGCTGGACACGACAAATCGCTGAATCGCTGATTCGATCACT 1200
QY 882 tccaatgatttcctaaatcctgcacccgtgtaagagcttgaaatgccttttactttga 941
Db 1201 TCTATCATCTCTCAAGGATTACATCTCTGCACAGGAGAAAGGAAATGTCCTGATGACT 1260
QY 942 agaaaaaggttttg 956
Db 1261 GGAGATGTCGCAGG 1275

RESULT 13

CELH28G03 44352 bp DNA INV 24-SEP-1999
LOCUS Caenorhabditis elegans cosmid H28G03.
DEFINITION AF098501
ACCESSION AF098501.2 GI:5921681
VERSION
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 44352)
The C. elegans Genome Sequencing Consortium, Washington University
Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
Hinxton, U.K., C.C.

TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)

ERRATUM: [[published erratum appears in Science 1999 Jan
1;283(5396):35]]
2 (bases 1 to 44352)

Wamsley, P. and Gibson, A.
The sequence of C. elegans cosmid H28G03

JOURNAL Unpublished (1999)
MEDLINE
REFERENCE 3 (bases 1 to 44352)

Waterston, R.
Direct Submission

JOURNAL Submitted (13-OCT-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS 4 (bases 1 to 44352)

Waterston, R.
Direct Submission

JOURNAL Submitted (24-SEP-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
e-mail: rwnematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' clone is T03G11, 4800 bp overlap; 3' clone is Y34BA, 6600 bp
overlap. Actual start of this clone is at base position 78 of
CELH28G03; actual end is at 37952 of CELH28G03

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source

Location/Qualifiers

1..44352
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="H28G03"
/chromosome="X"
6742..16476
/gene="H28G03.2"
join(6742,7098,7438..7536,7588..7643,12371..12518,
15241..15466,15634..15705,15761..16121,16167..16298,
16354..16476)
/gene="H28G03.2"
/note="coded for by C. elegans CDNA yk143c10.5; coded for
by C. elegans CDNA yk143c10.3; coded for by C. elegans
CDNA yk196a6.3; coded for by C. elegans CDNA yk196a6.5;
coded for by C. elegans CDNA yk255d2.3; coded for by C.
elegans CDNA yk255d2.5; coded for by C. elegans CDNA
yk413e1.5; coded for by C. elegans CDNA yk447a9.3; coded
for by C. elegans CDNA yk447a9.5; coded for by C. elegans
CDNA yk449a11.3; coded for by C. elegans CDNA yk449a11.5;
coded for by C. elegans CDNA yk467c2.3; coded for by C.
elegans CDNA yk467c2.5; coded for by C. elegans CDNA
yk470g12.3; coded for by C. elegans CDNA yk470g12.5; coded
for by C. elegans CDNA yk480a11.3; coded for by C. elegans
CDNA yk480a11.5; coded for by C. elegans CDNA yk483e7.3;
coded for by C. elegans CDNA yk483e7.5; coded for by C.
elegans CDNA yk492a8.3; coded for by C. elegans CDNA
yk492a8.5; coded for by C. elegans CDNA yk494b5.3"

/codon_start=1
/protein_id="AAC67404.1"
/db_xref="GI:3786418"
/translation="MQRDNVDSAGNPVNPVDLDQGEICESEFEATQAYLQDMEN
SAEIALERQRAEAVMATDDEDDGLNEFGRRNPRTPEPAEQGTCEIGFVN
EEFDVSDGELPGKRIGPKTPPEEPESPKQIEQTEQSGDELGTGDMPEPIIEP
KLKESREKREHATREHVRKRHRDDKRSVREVSKEITFPVAVATOKLSLEFVE
KRGLTIDDELQVVRSDIELDQAMGKIKKKEAMKQMEDMAWELEKIRIEMMP
RMDQVLKFDGKTVHIAQERPPMPMPFSSLSVSKYQPMHMGMPMPGMPMPF
MPPPIGMPMPPLGMPPHIGLGAAPYAPVPPSLGPPFPVPPSPPTSAAPPIV
PPPVQSTAQPPPSGTSEAKNMTGHLNPIAFSNAPGLKMGMDYQNPAPKPPAA
QONFQMRITNNLSMLTNALKQVSKAQNLSNLSSTSTAPKPPVPSLMSINIPGP
KAGSSGSON"
17763..18923
/gene="H28G03.1"
join(17763..18017,18067..18228,18281..18781,18831..18923)
/gene="H28G03.1"
/note="Contains similarity to Pfam domain: PF00076 (rrm),
Score=105.1, E-value=4.4e-28, N=2; coded for by C. elegans
CDNA yk171a8.5; coded for by C. elegans CDNA yk171a8.3;
coded for by C. elegans CDNA yk295h11.5"
/codon_start=1
/protein_id="AAC67401.1"
/db_xref="GI:3786415"

/translation="MNHECLLRPMHISLITILELPYLDWTFVKSKNNKLGVRSI
FKYALFOQFMGMGOYHDSPPOLKLFIGLSHDTDEOLGNVFSQWGPVDAIVI
EDPTKHSRGCFVTFFASIFAESAMNDRPHKLGKTVDSKRAIPREOMSSMIPPPF
ETDPAGCKLLNGITNGVSDLSURVTFGTLDQVEILGQRLGUGFVIYEDKESA
DRCIAHSGRHHVNERKIEVRVTKHPNGSTYWRKPSQSHSRDLFEQLSKDKG
DRKSTGNSSAADTPQNFDEDSNYGGTTTDEDCNCFEHEGSSSESSSTEQTLENEKEN
SD"
complement(21324..22960)

gene

CDS

gene

IMPORTANT: This sequence is not the entire insert of clone Y18D10A. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone F08A8 is at 152766 in this sequence. The true right end of clone B0019 is at 105 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence AL008866.

The end of this sequence (152766..152878) overlaps with the start of sequence Z99710.

For a graphical representation of this sequence and its analysis see: - <http://wormbase.sanger.ac.uk/per1/ace/elegans/sec/sequence?name=Y18D10A>.

FEATURES

source

Location/Qualifiers

1..152878

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="I"

/clone="Y18D10A"

complement(join(7920..8081,9752..9924,10879..11164,12369..12699,13906..14204,15129..16655,17748..18104,19860..19949,20008..20088,20709..20807,22205..23127,24486..24744,25724..25909,26020..26130,26236..26256))

/gene="Y18D10A.1"

complement(join(7920..8081,9752..9924,10879..11164,12369..12699,13906..14204,15129..16655,17748..18104,19860..19949,20008..20088,20709..20807,22205..23127,24486..24744,25724..25909,26020..26130,26236..26256))

/gene="Y18D10A.1"

/note="predicted using Genefinder"

cDNA EST YK541B5.5 comes from this gene"

/codon_start=1

/protein_id="CAA22308.1"

/db_xref="GI:3979929"

/db_xref="SPTREMBL:Q9XW25"

/translation="MDEDIAEACASGECANVRKNGLLMLQNKQRAAEQMDFSKVQA
ECEVLKMLVSDAAYRESVQREAEINAKAEASLQKQYDSFEYKKKSLAQEQA
DAEKKHQANTWNRKYSCALTELEANNIDIKSYESTVKLLETAGKRTSLDNO
VTLAKAIENTSKTKLDYKLSVQVVDALSDISNLSKQSKIDLVTFVRRODR
LFNFSQPNKSLSDSPISATAEASEDEPHLDVITPAPPICEOTLSDDEEEEEE
VPEADAEVDREIROLENELLRQVKEKTRRPDTAKMIELKAKSRAPVETKFK
KPEGIHKPTKXDSGLNSTKIHRNQFTSFDFIAQTAERKALLGAQAAGASE
PSSSIHGKKTVELTEEMKGVREKARMAEASAGRSQAPGAPAPAAASEL
QPPQDFGLMSDPSDSETEHQEKHIPAMVTRRSARLSALPVPKKASSSKMP
PPSPSPTRGRRRPTLSTMSPEAAAVTPAPGRPRSRSAKVSENPTEPSEA
PSAPVKRGRRSRSTMSITDSEPTSSSTAARKSKRAESEDDEEODLTKNKSPEK
PKPKSTTEVDGLKRLDRTATVITPGPPLTRKWRMRAPTAVTSKKE
KPKNAGSADSSINEEHEDETMILEQTLDLQOQPSQOEPRISSCELLEDPASEE
HSGTVSAPELTKNAPVPEASEASEAPPKIDIPQATPILALALALPVSPTALEP
PRAQENPTAELPTTSEISGRAPALPTSSQPTTSGSAAPPVDDLLSELGAKTTK
RKAAPPAVQKISSTQAPPTSVQAPPTSCSAAPPVDDLLSELGAKTTKTTK
MPVPDQKTISSAPPTISDSAPTSVHQPTPKPKQILNSKYGLDIDSDEDEEEEREG
MEIVEEBEPAPISDSLOASEPSTATVKPERVAVVKIFSPEDISTVEAPEASVP
PKAPATKIDNOLAQOASEPEPKARKLPIARKIPPKIKISLPAPSSSTSDDLMS
DILAGAKTTKPKAPRVAQTRTKNLAOKRASPTPAGTAPKPKQIKKSDS
VLPSSSSSTPPSPADSASTSSMMKGGGAIEMEAPCPGGAIKRQKPIGKEM
MVQNVKGGKYNKIKHLRQALDLKIPPEELKRPMEKGIKGLSDIPLTPSDAVDM
MEFLRETSAGNRCRLIRLFCHAVRFSESAENPDVTSPIYOKKSWIRRLFOILL
IFELCARSIGNECRLIRLFCHAVRFSESAENPDVTSPIYOKKSWIRRLFOILL
LNHNOIGKSTCYILMSDVAKYCRELVVEEMDKHMGSEFFLAFVLMHKDEQASAV
NMLLNKAFDVAVAPPTSDNDVTEACINAKQCIDDVDTSPSLKKSIFLAKTAASEVL
KNVFLRMDNPFSESHQFLKPKMEKEPRKRTAVSVPTVTKMTSEAMTKGASAV
KNAFLMDNPFSESHQFLKPKMEKEPRKRTAVSVPTVTKMTSEAMTKGASAV
QNLVDLITYTNFLPKPTTLPSS"

complement(join(26502..26573,27346..27405,27490..27580,27624..27788,28732..29006))

/gene="Y18D10A.2"

complement(join(26502..26573,27346..27405,27490..27580,27624..27788,28732..29006))

/gene="Y18D10A.2"

/note="predicted using Genefinder"

/codon_start=1

/protein_id="CAA22309.1"

/db_xref="GI:3979930"

/db_xref="SPTREMBL:Q9XW24"

/translation="MQMNPKNKPPISINFPLEMTSKFLFISISAVATASDLVEVFGI
SRCPDTSKFIHNLQVFPYQNYGNLSGLKLDHFHAVPTGGHVDGKYVNRCLHGALEC
ALNKQMSCKKHIKODMLVTAGCIOGTAYSAGLKCLPDTSEKLVQNCASEGEYVL
LNDENSRYNVAPHSAWLPMIQVNGERNRAREFLKDFSSVFLFEVFLQCLRSI
KN"

gene

complement(join(29879..30346,31220..31296,31346..31522,34557..34852,35323..35351))

/gene="Y18D10A.3"

complement(join(29879..30346,31220..31296,31346..31522,34557..34852,35323..35351))

CDS

/gene="Y18D10A.3"

/codon_start=1

/protein_id="CAA22319.1"

/db_xref="GI:3979940"

/db_xref="SPTREMBL:Q9XW15"

/translation="MDGSEKAIGLQFSLNAREWTKOLLATDEPFSSENVQVVR
FRITSEKRPKSVDKLYKTGEIIPARGKNKGVRSKMPISKSMELSLVKLQTPILQ
OFYKNGKSTWVHIISKTVSFQKLAQDDEQLFTKYGFQVQLMELAGLAAQA
IAAHPKSNVAVLCGPGNGGDFVCARHLQDQGFPTPSIVIPKESRNEMLKSLVQCE
TSSIPITATLPTNLQAFPLVDALFGFSFHPPTREPFTTEMLKTVRASGTHVFSIDVPS
GWDVILGAPSGNDDVIVHPSVISLTPKLCKMKNMTGPHFLGGREFVPKSLVDEHELLM
POYPGFEQIVKLED"

gene

join(35614..35976,36025..36138)

/gene="Y18D10A.4"

join(35614..35976,36025..36138)

CDS

/gene="Y18D10A.4"

/note="predicted using Genefinder"

/codon_start=1

/protein_id="CAA22310.1"

/db_xref="GI:3979931"

/db_xref="SPTREMBL:Q9XW23"

/translation="MTGERQTEAVSKRSNVWSERCGPPCLRLDGRILSVCSDDSS
HDGSGVHFFSLDCLSLUSTLSTAATVATATSSSSSKAFGVSLFLWSLLAWYV
VGVLVITGMDREKIVYVQFFHRLTPKRIENSAFSSFFCFLRPMIIGFTLSF"
join(39072..39164,40245..40513,41360..41936,44226..44375)

gene

/gene="Y18D10A.5"

join(39072..39164,40245..40513,41360..41936,44226..44375)

CDS

/gene="Y18D10A.5"

/note="predicted using Genefinder"

contains similarity to Pfam domain: PF00069 (Eukaryotic protein kinase domain), Score=322.1, E-value=2.1e-93, N=1
cDNA EST YK233B4.3 comes from this gene
cDNA EST YK233B4.5 comes from this gene
cDNA EST YK307B4.5 comes from this gene
cDNA EST YK451B9.5 comes from this gene
cDNA EST YK387B2.5 comes from this gene
cDNA EST YK337B9.5 comes from this gene
cDNA EST YK463B11.5 comes from this gene
cDNA EST YK507B2.5 comes from this gene
cDNA EST YK508B2.5 comes from this gene
cDNA EST YK520B5.5 comes from this gene
cDNA EST YK541B5.5 comes from this gene
cDNA EST YK590B9.5 comes from this gene
cDNA EST YK640B7.5 comes from this gene"

/codon_start=1

/protein_id="CAA22311.1"

/db_xref="GI:3979932"

/db_xref="SPTREMBL:Q9U2Q9"

/translation="MNKOLLSCLSKSQKQVTVMSVATDGVQDQVEISYDQKVI
GSFVVFVLALEYTNTNEMVAIKVQLQDKRFLQMRKLNHPNIVKLIFYSSGEK
KDELYLNLITLEYVETVYRVARHYSKQIQIPIYVLYMOLLRLSLAYIHSIGCHR
DLPKQNLITDPESGVLKCLDFGSKYLVRNEPNVSYIGSVYRPAFLGATNYNSI
DVMASGTMAELLGQIPFGDGVOLVEIILKVGTPTRQIQSMNPNYKEFKPQI
KAHPNKKVRVHTHPAEADLISKIIIEYPTSRPTQAACQHAFFDELNRPDLRLESGR
PULTEMGPMGTGEVSTTSGDVAGPSA"

gene

join(47525..47593,48112..48208,48610..48751,49616..50045,50501..50818,51553..51971,52106..52364,52847..53224)

/gene="Y18D10A.6"

CDS

join(47525..47593,48112..48208,48610..48751,49616..50045,50501..50818,51553..51971,52106..52364,52847..53224)

/gene="Y18D10A.6"

/note=contains similarity to Pfam domain: PF00999
(Sodium/hydrogen exchanger family), Score=354.0,
E-value=5.3e-103, N=1
/codon_start=1
/protein_id="CAA22320.1"
/db_xref="GI:3979941"
/db_xref="SPTREMBL:O9XW14"
/translation="MMLSEVQNRNNEELFTFNSLKIANGVWOLWESTNQCTA
TSGIMFAKLTALSTICFLLFQTVDSDDSSASVGSVAKSDETVAVNKTVL
GEAIDANVLSHGNAIVGNVSEKKSIALFELFVLMILAVHMLVLSKIHMP
ESLAIYVALAGLSILSYRRNDSIEALSFDVDFVLFPPIFENAYNNLKGYPFN
FVPLIFALGTTISLVAGGLYILGALGILFETFEFCFAAMISADVPGTIAL
FOAKVSEGLMLVFEENMLNDASVILAAITLKHAKSPNSPASEIISTASVTE
MFFFSACGLGLGLLGLFELKHVDLRKTSLEFALLIFSYIPGFAELDLSGMAI
LFCGISMSQFTRHNSYQAQITFRFTITSEFVATTFYIGNAFTIKLNFAPWL
FWSVGLILGRACVNPFLAYNQKDOVLSMKNNQIMFSGMRGACVFLAYLMDL
DKKKSILITLVLTFLFTIPGLSGALPFTFYINQYENPQRRRTPRNKSTGN
SALMKSSQEMSFSTGSDWGPKSAALDTSAGRIQMLQRFVKRTAERLENKDLA
ILKRALASDQSDSDVEFGGGGCGVGGGRKKDDVTPTRGRSGRNSSDVISAGG
GUGSGHLLISGSDSNFF"

[illegible]

RESULT	15				
LOCUS	CEY18D10	247332 bp	DNA	HTG	07-SEP-1999
DEFINITION	Caenorhabditis elegans chromosome I clone Y18D10, *** SEQUENCING IN PROGRESS ***, in unordered pieces.				
ACCESSION	AL008871				
VERSION	AL008871.1	GI:3900862			

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT
ORIGIN

```

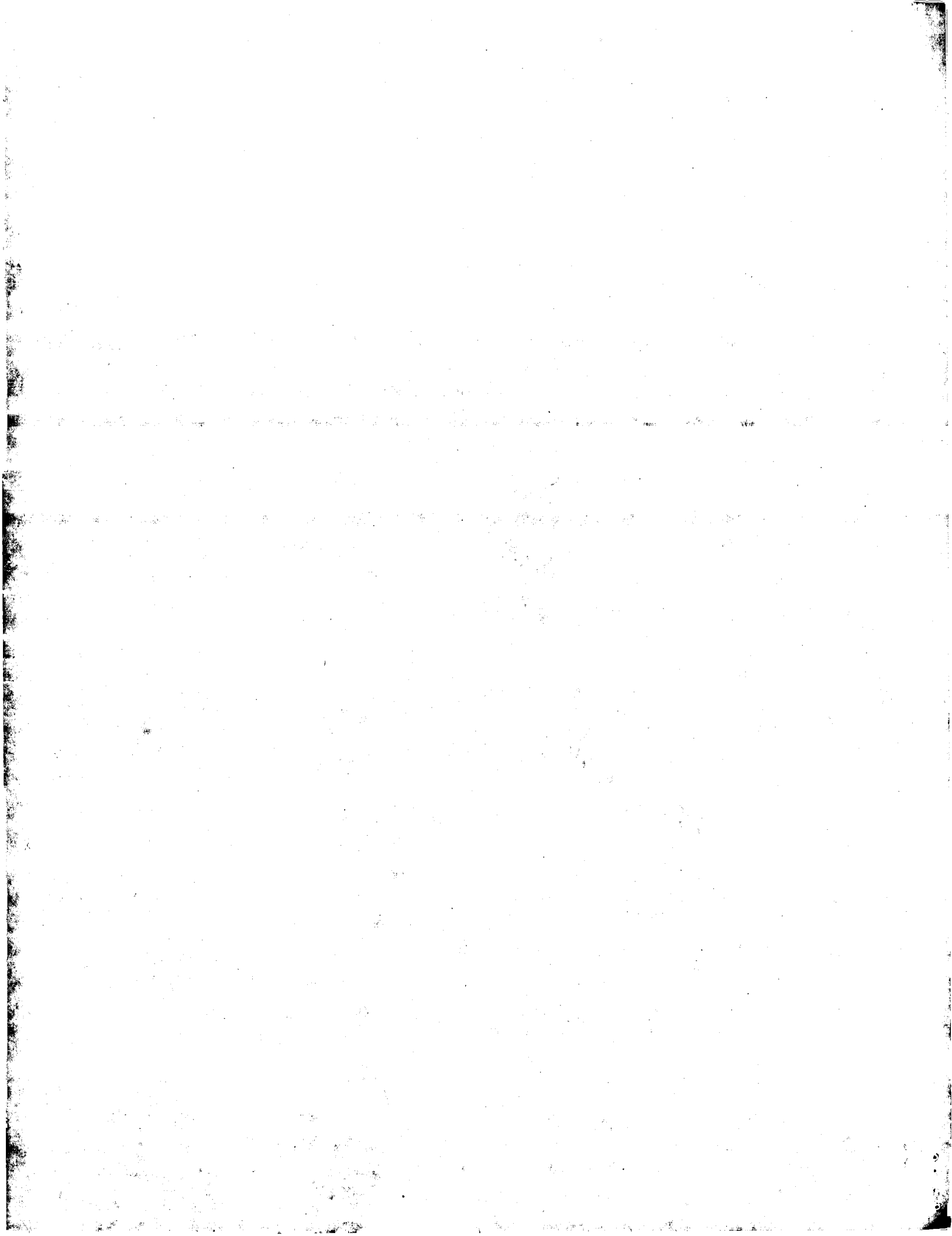
HTG; HTGS_PHASE1.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 247332)
Harris, B.
Direct Submission
Submitted (07-SEP-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
On Nov 21, 1998 this sequence version replaced gi:3973430.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

Location/Qualifiers
1. 247332
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="I"
/clone="Y18D10"

79342 a 43026 c 42968 g 79596 t 2400 others

[illegible]

Search completed: June 29, 2001, 12:01:42
Job time: 8915 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 11:49:30 ; Search time 162.85 Seconds
(without alignments)
3759.313 Million cell updates/sec

Title: US-09-668-788-3

Perfect score: 975

Sequence: 1 atgtgtactcaataaaaa.....gtaaaacgtgatactccag 975

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	883.2	90.6	12658	18 AAV74438	Staphylococcus aur
c 2	437	44.8	2600	AA080512	MurE gene. Staphy
c 3	182.2	18.7	936	22 AAF58252	Oligonucleotide D1
c 4	182.2	18.7	936	22 AAF58254	Oligonucleotide D1
c 5	182.2	18.7	936	22 AAF58257	Oligonucleotide D1
c 6	182.2	18.7	936	22 AAF58259	Oligonucleotide D2
c 7	182.2	18.7	936	22 AAF58262	Oligonucleotide D2
c 8	182.2	18.7	936	22 AAF58255	Oligonucleotide D1
c 9	175.2	18.0	936	22 AAF58252	Oligonucleotide D1
c 10	175.2	18.0	936	22 AAF58254	Oligonucleotide D1
c 11	175.2	18.0	936	22 AAF58257	Oligonucleotide D1

12	175.2	18.0	936	22	AAF58259	Oligonucleotide D2
13	175.2	18.0	936	22	AAF58262	Oligonucleotide D2
14	175.2	18.0	938	22	AAF58255	Oligonucleotide D1
15	118	12.1	1149	20	AAF23386	B. subtilis glycos
c 16	66	6.8	244	22	AAF58238	Oligonucleotide D1
17	65.8	6.7	244	22	AAF58238	Oligonucleotide D1
c 18	51.2	5.3	1764	21	AAC47190	Arabidopsis thalia
c 19	47.2	4.8	580073	18	AAT58840	Mycoplasma genital
c 20	47	4.8	1265	19	AAV12734	Monogalactosylac
c 21	47	4.8	11011	21	AAC68252	B. burgdorferi tox
22	46.4	4.8	8920	15	AAQ62924	Carbamoyl-phosphat
23	45.8	4.7	11922	21	AAV70187	Plasmodium falcipa
24	44.6	4.6	163319	21	AAF22306	Arabidopsis thalia
25	43.8	4.5	19124	18	AAT72882	Plasmodium var-7 g
26	43.8	4.5	19124	21	AAZ98287	Plasmodium var-7 p
27	43.4	4.5	9789	17	AAT41852	CDNA encoding Plas
28	43.2	4.4	32768	20	AAI13037	Enterococcus faeca
c 29	42.8	4.4	549	21	AAC95242	Cat flea head and
30	42.8	4.4	1431	21	AAZ37082	DNA sequence encod
31	42.8	4.4	1671	13	AAQ24134	50 kD subunit of S
c 32	42.6	4.4	116277	20	AAZ20249	Borrelia burgdorfe
c 33	42.4	4.3	1416	21	AAV70135	Plasmodium falcipa
c 34	42	4.3	3366	21	AAV70175	Plasmodium falcipa
c 35	42	4.3	4590	7	AAAG0472	Sequence encoding
c 36	42	4.3	19124	18	AAV72882	Plasmodium var-7 g
c 37	42	4.3	19124	21	AAZ98287	Plasmodium var-7 p
c 38	41.8	4.3	2486	21	AAV7037	Nucleotide sequenc
39	41.4	4.2	6265	20	AAV08523	NBP46 (root lectin
40	41	4.2	2058	19	AAV07560	Neocallimastix pat
41	41	4.2	2058	20	AAZ11460	N. patriciarum xyl
42	41	4.2	2058	22	AAC66514	N. patriciarum end
c 43	40.8	4.2	3369	21	AAV70230	Plasmodium falcipa
44	40.6	4.2	4317	21	AAV70204	Plasmodium falcipa
45	40.6	4.2	7143	21	AAV70250	Plasmodium falcipa

ALIGNMENTS

RESULT 1

ID AAV74438 standard; DNA; 12658 BP.

XX AAV74438;

XX AC

DT 16-MAR-1999 (first entry)

XX Staphylococcus aureus contig SEQ ID #127.

Computer readable medium; vaccine; S.aureus infection; Immunodetection; cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy; skin infection; surgical wound infection; scalded skin syndrome; toxic shock syndrome; ds.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT misc_feature

FT 181..240

FT /*tag= a

FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence".

FT misc_feature

FT 1981..2040

FT /*tag= b

FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence".

FT misc_feature

FT 3781..3840

FT /*tag= c

FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence".

are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

5581..5640
/*tag= d
/note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

7381..7440
/*tag= e
/note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

9181..9240
/*tag= f
/note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

10981..11040
/*tag= g
/note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

EP786519-A2.
30-JUL-1997.
07-JAN-1997; 97EP-0100117.
05-JAN-1996; 96US-0009861.
(HUMA-) HUMAN GENOME SCI INC.
Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
Rosen CA;
WPI; 1997-374922/35.
Polynucleotide(s) and proteins derived from Staphylococcus aureus
stored on computer readable medium and used in the production of
anti-S.aureus vaccines
Claim 1; Page 692-699; 3271pp; English.
This sequence represents one of 5191 Staphylococcus aureus DNA sequences
of the invention. The DNA sequences are recorded on a computer readable
medium, preferably selected from a floppy or hard disk, random access
memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
the S.aureus DNA sequences allows putative functions to be assigned so
that protein-encoding or regulatory regions of commercial, therapeutic or
industrial importance can be obtained. Specifically, sequences which are
likely to encode antigens have been identified and these polypeptides can
be used in a vaccine composition against S.aureus infection. The
polypeptides can also be used in a kit for the immunodetection of
S.aureus in a sample. S.aureus is implicated in numerous human diseases,
including cellulitis, eyelid infections, food poisoning, osteomyelitis,
skin and surgical wound infections, scalded skin syndrome, toxic shock
syndrome, etc. Organisms transformed with the DNA sequences can be used
for recombinant production of the polypeptides. The new DNA sequences
(and their fragments) are useful as primers or probes for isolating
homologues of any of the S.aureus DNA sequences contained on the
computer readable medium.

Sequence 12658 BP; 4296 A; 1711 C; 2283 G; 3942 T; 426 other;

Query Match 90.6%; Score 883.2; DB 18; Length 12658;
Best Local Similarity 92.4%; Pred. No. 1.5e-183;
Matches 902; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 1 atggttactcaataataaaagatattgatttattactggtcattcggttaacggtcatatg 60
|||||
Db 5652 ATGTTACTCAANN 5593
QY 61 caagtacacagagatogttaaatacaactaatgatataatgaatcatttaagcgtc 120
|||||
Db 5592 NNNNNNNNNNNAGTATCGTTAAATCAACTTAATGATGAATCTAGACCATTTAAGCGTC 5533
QY 121 attgagcacgattattatagaagtcacccaattttgactcttattgttaaaaaatgg 180
|||||
Db 5532 ATTGAGCAGCAATTTATTATGGAAGCCTATCCCAATTTGACTTCTTATTTGTAATAATGG 5473
QY 181 tataatcagctcttaataatttttagaaatgtacaaaggggtttttattacagcgccca 240
|||||
Db 5472 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTATTATACAGCGCCCA 5413
QY 241 gataaactagacaatgttttttaacaatactatgacttaataagtttaataattattg 300
|||||
Db 5412 GATAAAGTACAAATGTTTTCACAAATACATGAGCTTAATAAGTTAATTAATTTATTG 5353
QY 301 ataaaagaaagccagatttattattatttaacgttttcctacaccagttatgctgacta 360
|||||
Db 5352 ATAAAGAAAGCCAGATTTAATATTATTAAGTTTCTCTACACAGTTATGTCGTACTA 5293
QY 361 actgagcaatttaacatttaattccagttgtctacagtgatgacagactatcgcttaacat 420
|||||
Db 5292 ACTGAGCAATTTAATTAATTTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 5233
QY 421 aaaaactggtatgcgcgtatttcaacaagatattatggtgcaacaagaagaaacacaa 480
|||||
Db 5232 AAAAAGTGGATTACGCCGTATTTCACAAAGATATTATGTTGGCAACAAAGAAACGAAACAA 5173
QY 481 gacttcacagcgttagtgattgattccttcaacagtttaagtgacaggttattcctattgat 540
|||||
Db 5172 GACTTCATAGAGGTAGGTATTGATCTCTTCAACAGTTAAAGTGACAGGTATTCTTATTGAT 5113
QY 541 acaaaatttgaaacgcctatttaatacaagaagcgtggtttaaataagacacaaactatgacca 600
|||||
Db 5112 AACAAATTTGAAACGCCCTATTAAATCAAAAGCAGTGGTTAATAGACAAACACTTAGATCCA 5053
QY 601 gataagcaaacatttattatgctcagctggtgcttctggttatctaaaggttttgacacg 660
|||||
Db 5052 GATAGCAAACTATTTTAAATGTCACGTGGTCATTTGGTGATCTTAAAGTTTTCACAGC 4993
QY 661 atgattactgatataattagcgaaaagtgcaaatgcacagtgattgattggttgtaag 720
|||||
Db 4992 ATGATTACTGATATATTAGCGAAAAGTGCAAATGACAAAGTAGTATTGATTTTGTGTAAG 4933
QY 721 agcaagagctaaagcgtttctttaacagctaaagttaataacaggaatgtattgatt 780
|||||
Db 4932 AGCAAGAGCTAAAGCGTTCTTTAAACAGCTAAGTTTAAATCTAACGAGNAATGATTGATT 4873
QY 781 ctagggttatacaaacacacatgaatgaatggatggatcgaatcaagtcaacttatgattacgaaa 840
|||||
Db 4872 CTAGGTTATACCAACACATGAATGAATGGATGGCATCAAGTCAAGTCAACTTATGATTACGAAA 4813
QY 841 cctggtggtatcaacaataactgaaggtttcgcccggttgattcccaatgatttctctaaat 900
|||||
Db 4812 CCTGTTGGTATCACAAATACTGAAGGTTTCGCCCGGTTGTTATTCCAATGATTTCGTAAT 4753
QY 901 cctgcacctggtcaagagccttaaaaatgcttttactttgaagaaaggttttggtaaa 960
|||||
Db 4752 CTGTCACCTGGTCAAGAGCTTGAATAATGCCCTTACTTTTGAAGAAAGAGGTTTGGTAAA 4693
QY 961 a-cgctgatactccag 975
|||||
Db 4692 ATCGCTGATACCTCCAG 4677

RESULT 2
AAX08512/C
ID AAX08512 standard; DNA; 2600 BP.

XX AAX08512;
AC
XX 19-JUL-1999 (first entry)
DT
XX Mure gene.
DE
XX Mure; Staphylococcus aureus; infection; treatment; diagnosis;
KW primer; osteomyelitis; septic arthritis; septic thrombophlebitis;
KW PCR; scalded skin syndrome; antibody; antibiotic; ds.
XX
OS Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FT 93..2473
CDS /*tag= a
FT /product= "UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:
FT lysine ligase"
XX
XX EP905248-A1.
PN
XX 31-MAR-1999.
PD
XX 21-SEP-1998; 98EP-0307624.
PF
XX 10-DEC-1997; 97US-0988251.
PR
XX 25-SEP-1997; 97US-0059956.
PR
XX (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PA
XX Shilling LK, Wallis NG, Zabierowski S;
PI
XX WPI: 1999-206775/18.
DR P-PSDB; AAN85670.
DR
XX New Mure gene from Staphylococcus aureus useful for diagnosing and
PT treating bacterial diseases such as osteomyelitis and toxic shock
PT syndrome
XX
XX Claim 6; Page 5-8; 32pp; English.
PS
XX Staphylococcus aureus can cause both invasive and toxigenic
CC infections which lead to diseases such as osteomyelitis, septic
CC arthritis, septic thrombophlebitis, scalded skin syndrome and toxic
CC shock syndrome. Antagonists of Mure such as antibodies may be used
CC to treat individuals with these diseases and as a prophylactic
CC against them by bathing wounds and coating implanted devices.
CC Antagonists of Mure may also be used to treat diseases caused by
CC Helicobacter pylori infection such as stomach ulcers and
CC gastrointestinal carcinoma. This method of treatment of bacterial
CC infection is advantageous since many strains of Staphylococcus
CC aureus are now resistant to some or all of the standard antibiotics.
CC Mure is an important enzyme in bacterial metabolism and so the
CC identification of such an enzyme in Staphylococcus aureus can be
CC used to effectively diagnose and treat diseases caused by this
CC pathogen. Detection of mure, and hence a bacterial infection,
CC can be done by reverse transcriptase polymerase chain reaction
CC (RT-PCR), preferably using the primers described in AAX08514, AAX08515.
XX
XX Sequence 2600 BP; 893 A; 370 C; 493 G; 844 T; 0 other;
SQ

Query Match 44.8%; Score 437; DB 20; Length 2600;
Best Local Similarity 100.0%; Pred. No. 1.8e-86;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtgtactcaaaataaaagatattgattacttggtcattcggtgaacggtcatatg 60
DB ATGTTACTCAAAATAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498
QY 61 caagttacacagagtatcgttaacttaacttaataatgatgaatgaaccatttaacgctc 120
|||||

Db 497 CAAAGTTACAGAGTATCGTTAATCAACTTAATGATGATGAATCTAGACCATTTAAGCGTC 438
QY 121 attgagcacgatttatttgaaagctcatccaattttgacttctctattgtaaaaaatgg 180
|||||
Db 437 ATTGACGACGATTATTATTATGGAAGCTCATCCAATTTTGACTTCTATTGTAATAATGG 378
QY 181 tatatcaatagcttttaaatatttttagaataatgtacaaaggggtttattacagcgccca 240
|||||
Db 377 TATATCAATAGCTTTAAATATTTTAGAAATATGTAACAAGGGTTTATTATACAGCCGCCA 318
QY 241 gataactagacaaatggtttttacaaataactatggacttaataaagtttaattatttg 300
|||||
Db 317 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTATTG 258
QY 301 ataaaagaaagccagatttaataattattaaagtttccctacaccagttatgctggtacta 360
|||||
Db 257 ATAAAGAAAGCCAGATTAAATATTATTAACGTTTCCCTACACCAGTTATGTCGGTACTA 198
QY 361 actgagcaatttaacatttaattccagttgctacagtgatcacagactatcgcttacct 420
|||||
Db 197 ACTGACCAATTTAACATTAAATATTCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 138
QY 421 aaaaactggattacgcc 437
Db 137 AAAAAGTGGATTACGCC 121
RESULT 3
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
XX WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

```
Query Match      18.7%; Score 182.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 4.8e-31;
Matches 7; Conservative 532; Mismatches 240; Indels 0; Gaps 0;

Qy 5 ttactcaaaataaaagattgtattactgctcattcgttaacggtcatatgcaag 64
Db 779 WWWWWW... 720
Qy 65 ttacacagagatcgttaatacaacttaataagtttaacagcgcgcagata 124
Db 719 WWWWWW... 660
Qy 125 agcagatttattgaagctcatcaatttgcattcttattgtaaaatgggata 184
Db 659 WWWWWW... 600
Qy 185 tcaatagctttaaataatttagaataatgtacaaaggtttattacagcgcgcagata 244
Db 599 WWWWWW... 540
Qy 245 aactagacaaatgttttacaataactatgagacttaataagtttaatttgataa 304
Db 539 WWWWWW... 480
Qy 305 aagaaagcagatttaataatttaacgcttctcaccagttatgtcggtactaactg 364
Db 479 WWWWWW... 420
Qy 365 agcaatttaacatttaataatccagttgctacagtgatgacagactatcgttatacaaaa 424
Db 419 WWWWWW... 360
Qy 425 actgattacgcggtattcaacagatattatgtgcaacaaagaaacgaacagact 484
Db 359 WWWWWW... 300
Qy 485 tcatagacagtaggtattgatactcctcaacagtttaagtgacaggtattccttataca 544
Db 299 WWWWWW... 240
Qy 545 aattgaaacgcctatttaacagcagtaagtttaataacagagaatgtattgattcta 604
Db 239 WWWWWW... 180
Qy 605 agcaactatttaagtcagctgctgattgtgtatctaaaggttttgacacagta 664
Db 179 WWWWWW... 120
Qy 665 ttactgatattagcgaagtgcaaatgacaaagtaggtattgtgtgtaagagca 724
Db 119 WWWWWW... 60
Qy 725 aagactaaagcgtctttaaagcagtaagtttaataacagagaatgtattgattcta 783
Db 59 WWWWWW... 1
```

```
RESULT 4
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
```

```
XX 01-FEB-2001.
PD
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
SQ

Query Match      18.7%; Score 182.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 4.8e-31;
Matches 7; Conservative 532; Mismatches 240; Indels 0; Gaps 0;

Qy 5 ttactcaaaataaaagattgtattactgctcattcgttaacggtcatatgcaag 64
Db 779 WWWWWW... 720
Qy 65 ttacacagagatcgtttaaacttaataagtttaacagcgcgcagata 124
Db 719 WWWWWW... 660
Qy 125 agcagatttattgaagcagctcatcaatttgcattcttattgtaaaatgggata 184
Db 659 WWWWWW... 600
Qy 185 tcaatagctttaaataatttagaataatgtacaaaggtttattacagcgcgcagata 244
Db 599 WWWWWW... 540
Qy 245 aactagacaaatgttttacaataactatgagacttaataagtttaatttgataa 304
Db 539 WWWWWW... 480
Qy 305 aagaaagcagatttaataatttaacgcttctcaccagttatgtcggtactaactg 364
Db 479 WWWWWW... 420
Qy 365 agcaatttaacatttaataatccagttgctacagtgatgacagactatcgttatacaaaa 424
Db 419 WWWWWW... 360
Qy 425 actgattacgcggtattcaacagatattatgtgcaacaaagaaacgaacagact 484
Db 359 WWWWWW... 300
Qy 485 tcatagacagtaggtattgatactcctcaacagtttaagtgacaggtattccttataca 544
Db 299 WWWWWW... 240
Qy 545 aattgaaacgcctatttaacagcagtaagtttaataacagagaatgtattgattcta 604
Db 239 WWWWWW... 180
```



```
Db 69 ..... 128
QY 125 agcagatttattgaagctcaatcttgaacttcttattgtataaaatggtata 184
Db 129 ..... 188
QY 185 tcaatagcttttaattttgaagaataatgtacaaagggtttttattacagcgccagata 244
Db 189 ..... 248
QY 245 aactagacaaatgtttttacaaatactatggacttaataagttattattattgataa 304
Db 249 ..... 308
QY 305 aagaaaagcagatttaattatttaacggttccacacaggttatgtcggtactaaactg 364
Db 309 ..... 368
QY 365 agcaatttaacatttaattccagttgtcacagtgatgacagactatcgcttacataaaa 424
Db 369 ..... 428
QY 425 actgattacgcgtattcaacagatattattgtggtatctaaaggaaacaaagact 484
Db 429 ..... 488
QY 485 tcatagacgtaggtattgatcttcaacagtttaagtgacaggtattcctattgataaa 544
Db 489 ..... 548
QY 545 aattgaacgccttatttaacaaagcagtggttaataagacaaacttagatccagata 604
Db 549 ..... 608
QY 605 agcaactatttttaagtcagctggtgcatgtgtatctaaagggttttgacacagatga 664
Db 609 ..... 668
QY 665 ttactgatattagcgaaaagtgcaaatgcacagtagttatgttggtaagagca 724
Db 669 ..... 728
```

RESULT 11

AAF58257 standard; DNA; 936 BP.

XX AAF58257;

AC AAF58257;

DT 24-APR-2001 (first entry)

XX Oligonucleotide D1954.

DE Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

OS WO200107665-A2.

PN 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

PD 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

PR (CLIN-) CLINICAL MICRO SENSORS INC.

XX PA

XX

PI Umek RM;

XX WPT; 2001-159728/16.

XX

PT Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PT a single surface

XX

PS Example 6; Page 127; 159pp; English.

XX

CC The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

XX

SQ Sequence 936 BP; 5 A; 142 C; 7 G; 776 other;

XX

Query Match 18.0%; Score 175.2; DB 22; Length 936;

Best Local Similarity 0.4%; Pred. NO. 1.6e-29;

Matches 3; Conservative 532; Mismatches 245; Indels 0; Gaps 0;

XX

QY 5 ttactcaaaataaagaatattgattactggtcgtacaggttcataatgcaag 64

Db 9 68

QY 65 ttacacagagatcgttaatacacttaataatgataatgacacatttaagcgtcattg 124

Db 69 128

QY 125 agcagatttattgaagctcaatcttgaacttcttattgtataaaatggtata 184

Db 129 188

QY 185 tcaatagcttttaattttgaagaataatgtacaaagggtttttattacagcgccagata 244

Db 189 248

QY 245 aactagacaaatgtttttacaaatactatggacttaataagttattattgataa 304

Db 249 308

QY 305 aagaaaagcagatttaattatttaacggttccacacaggttatgtcggtactaaactg 364

Db 309 368

QY 365 agcaatttaacatttaattccagttgtcacagtgatgacagactatcgcttacataaaa 424

Db 369 428

QY 425 actgattacgcgtattcaacagatattattgtggtatctaaaggaaacaaagact 484

Db 429 488

QY 485 tcatagacgtaggtattgatcttcaacagtttaagtgacaggtattcctattgataaa 544

Db 489 548

QY 545 aattgaacgccttatttaacaaagcagtggttaataagacaaacttagatccagata 604

Db 549 608

QY 605 agcaactatttttaagtcagctggtgcatgtgtatctaaagggttttgacacagatga 664

Db 609 668

QY 665 ttactgatattagcgaaaagtgcaaatgcacagtagttatgttggtaagagca 724

Db 669 728

QY 485 tcataagcgttaggtattatctcaacagttaaagtacaggtattcctctattgataca 544
 Db 489 wwwcwww 548
 QY 545 aatttgaaacgcctattatcaaaagcagtggttaataagacacaaacttagatcagata 604
 Db 549 wwwcwww 608
 QY 605 agcaaaactatttaagtgcagctggtgatttggtgtatctaaagggttttgacagatga 664
 Db 609 wwwcwww 668
 QY 665 ttactgatatattagcgaaagtgcaaatgcacagtagttatgatttggtgaagca 724
 Db 669 wwwcwww 728
 QY 725 aagagctaaagcgttctttaacagctaaagtttaataattacagagaatgtattgtattctag 784
 Db 729 wwwcwww 788

RESULT 15

AAZ23386
 ID AAZ23386 standard; DNA; 1149 BP.

AC AAZ23386;

DT 13-DEC-1999 (first entry)

DE B. subtilis glycosyl transferase catalytic domain DNA.

KW Catalytic domain; glycosyl transferase; processive activity; detergent;
 KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
 KW food emulsifier; polymer modifier; glycosyl transferase; ss.

XX Bacillus subtilis.

XX DE19819958-A1.

XX 30-SEP-1999.

XX 05-MAY-1998; 98DE-1019958.

XX 25-MAR-1998; 98DE-1013017.

XX (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.

XX (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.

XX Wolter FP, Jorasch P, Heinz E, Zaehring U;

XX WPI: 1999-552364/47.

XX P-PSDB; AAY33444.

XX New glycosyl transferase protein, useful for producing glycolipids

XX Example 1; Page 6; 8pp; German.

XX This invention describes a novel protein (I) comprising identical or
 CC different catalytic domains of glycosyl transferases and which has
 CC processive activity (i.e. builds up oligosaccharide glycolipids by
 CC successive addition of glucose units). (I) can be used to produce
 CC 3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
 CC modifiers or detergents. This sequence encodes a Bacillus subtilis
 CC glycosyl transferase catalytic domain which is used in the method of
 CC the invention.

XX Sequence 1149 BP; 380 A; 217 C; 244 G; 308 T; 0 other;

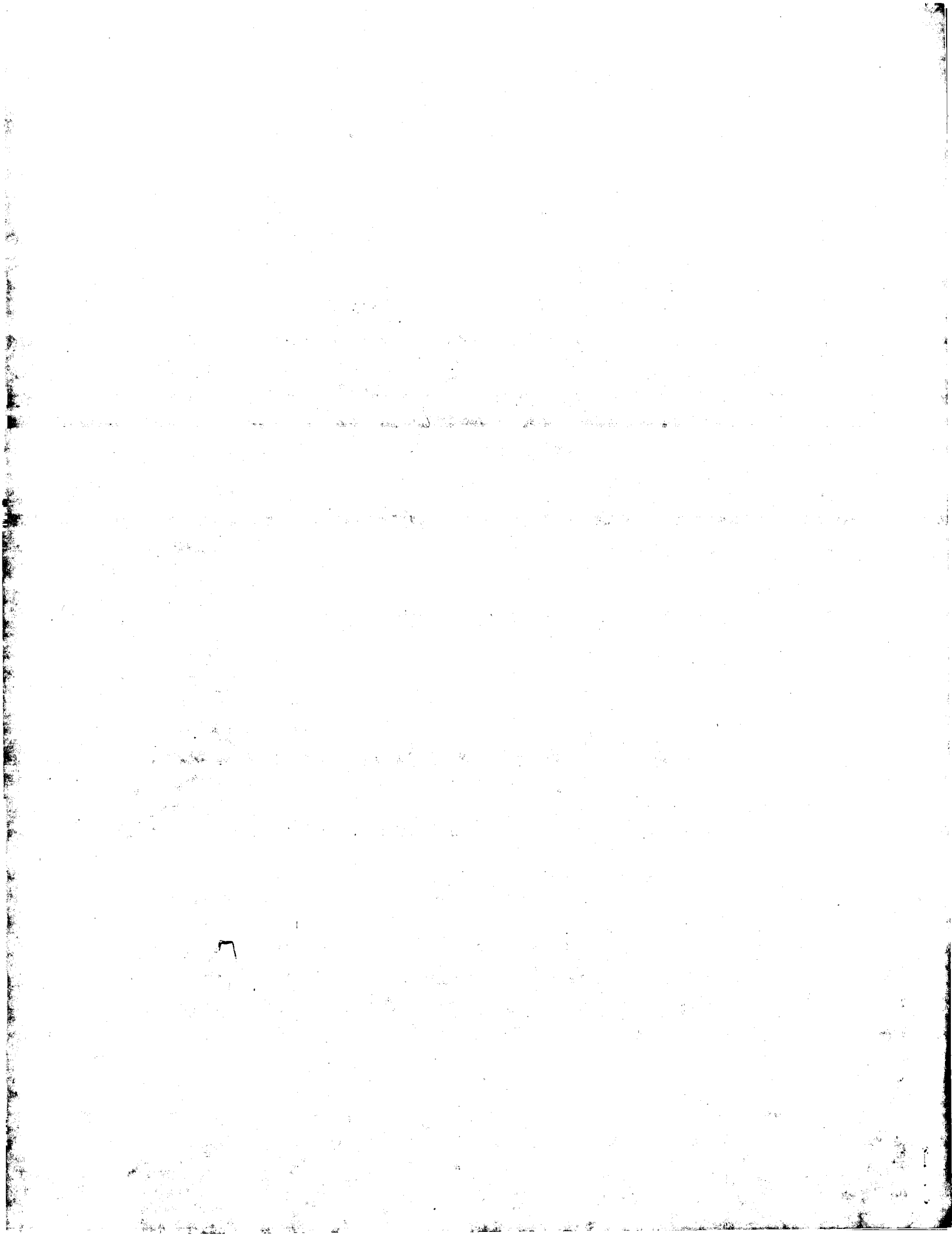
XX Query Match 12.1%; Score 118; DB 20; Length 1149;

XX Best Local Similarity 48.5%; Pred. No. 4.7e-17;

XX Matches 458; Conservative 0; Mismatches 465; Indels 21; Gaps 4;

QY 13 aataaaagatattgattattactgctcattcgcgttaacggtcatatgcaagttacacag 72
 Db 10 aataaaagatattattttgactgcaaatcagaaatgacatgacaggtgacgcaaa 69
 QY 73 agtatcgtaatacaacttaataatgatatgaatagaccatttaacgctcattgagcagat 132
 Db 70 acatttatgacaatgtgacggtcgttcagcttccagatgtaacagttctt-----aat 123
 QY 133 ttatttatggagctcatccaaattttgacttctattgtgtaaaaaatggtatcatatagc 192
 Db 124 ttgacaaagtgtaaatcccgattgtttcagaggttaactcaacttattttaaagc 183
 QY 193 tttaaatatttttagaaatatgtacaaagggttttatt-----acagccgcacagat 243
 Db 184 ttctcaatcgggaaacagttttatcgtttgtttattacggaggtgacaaaatcataat 243
 QY 244 aaactagacaaatgtttttacaaatactatgacttaataagtttaattatttata 303
 Db 244 aaactgtaattcaattatttactttaaaatgggtaaaaagattggcgacttgcgat 303
 QY 304 aaagaaaagccagatttaataattatttaacgttttctacacacagttatgcgttactaact 363
 Db 304 gaactcagccgataattatttaataattcattccgatgactcgtgcggaatacaga 363
 QY 364 gagcaatttaacatttaattccagttgtctacagtgatgacagactatcgtttacataaa 423
 Db 364 cgcggaactggaagagtcattcctacaccttaacgtttatgactgtttttgtcttcaaaa 423
 QY 424 aactgattacgcgtattcaacaagatatattatgtggcaacaaaagaacaaacagac 483
 Db 424 attgggttcagaaaacgttgataaaattattatgtggcagagattacgtgtaagaaaaa 483
 QY 484 ttcatagacgtaggtattgactccttcaacaggttaaaagtgacaggtattccttattgataac 543
 Db 484 ctgctggagatcggcactcattcccaagcaatgtaaaaaacacaggaattccaatcagccg 543
 QY 544 aaattgaaacgcctatttaatacaaaagcagtggtttaaataagacaacaaactagatccagat 603
 Db 544 caatttgaaatccatgcctgttgcccgatataataaaagtaacaattctttcaccaaac 603
 QY 604 aagcaaaactattttaaactcagctggtcatttgggtgtatctataaaggttttgacacgatg 663
 Db 604 aaaaaagtctctgtatcatgctgcaggtgctcagcgtgtattaaagaacgtaaaagagctg 663
 QY 664 attactgatatattagcgaagaagtgcacaaatgcacagtagttatgtattgtgtaagagc 723
 Db 664 tgc---gaaaacctgtcaaggatgaccaaagtgcaagtagttgtcgtgctgcggaataat 720
 QY 724 aaagagctaaagcgttcttttaacagctaagtttaaatgaagagagaaatgtatttga---tt 780
 Db 721 acggcttttaaaagaatctttgagtgctgcttgaaagcggaaaatggtgacaaaataaagt 780
 QY 781 ctaggttatacaaacacacatgaatgaatggatggcatcaagtcaactattattacgaaa 840
 Db 781 ctgggctatgtggcgcgtatgtagctatttccgatacagagattgcatgattaccaga 840
 QY 841 cctgtgtgtatcacataactgaagtttgcgcgttgcctgttattcccaatgatttccataat 900
 Db 841 ccgcgcgcgtattctttgacagaagccacagccattggagtcctgtcattctgtacaaa 900
 QY 901 cctgcacctgttcaagagcttgaataatgccttttactttgaaga 944
 Db 901 ccctgctcgtgcgcaagaaaaaagaataatgcaacttctttgaaga 944

Search completed: June 29, 2001, 11:49:52
 Job time: 5214 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 10:55:47 ; Search time 1964.27 Seconds
(without alignments)
4692.090 Million cell updates/sec

Title: US-09-668-788-3
Perfect score: 975
Sequence: 1 atggttactcaataaaaa.....gtaaaacgctgatactccag 975

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[illegible]

[illegible]

RESULT	4	
CNS00KB5	1101 bp	DNA
LOCUS		GSS
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR39P05 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL077453.1	GI:4956930
VERSION		GSS.
KEYWORDS	fruit fly.	
SOURCE	Drosophila melanogaster	
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 1101)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequences : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genosco.cns.fr)	
COMMENT	- Web : www.genoscope.cns.fr Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/Drosophila_bac.htm .	

Query Match	6.1%	Score 59.4	DB 219	Length 1101
Best Local Similarity	38.7%	Pred. No. 0.0032		
Matches	232	Conservative	340	Indels 0
Qy 172	aaaaaaatggtatcaatagctttaaataatatttttagaataatgatacaaggggtttttattac	231		
Db 143	AAAAAAAAAATAATAATAATATAAAAAATATATAAAAAATATTTAAATATATAAAAAATAA	202		
Qy 232	agcgcgcagataaactagacaaatgtttttcacaaatactactgacttaataagtttaatt	291		
Db 203	ATWATTAATAAATTAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	262		
Qy 292	aattattgataaagaagaacccgagtttaataatattataacagtttctctacacccagttatg	351		
Db 263	AAAAATAAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA	322		
Qy 352	cggtactaaactgagcaatttaacataaattccagttgctacagtgatgacagactat	411		
Db 323	AAANANNNNNNAATAAAAAAANAATATANANAATTTATAATATTTTTTTTATTTTTTT	382		
Qy 412	cgcttacataaaaactggattcacgcggtattccaagaagatatattgtggcaacaaagaa	471		
Db 383	ATTTTATATTTTAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA	442		
Qy 472	acgaacaagaactcatagacgtaggtattgatcccttcacacagtttaagtgacaggtatt	531		
Db 443	AWATTAATAAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA	502		
Qy 532	octattgatacaaaatgtgaacgccttataatcaaaagcagtggttaataagacacacac	591		
Db 503	AAATTTTTTTTTTAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA	562		
Qy 592	ttagatccagataagaacactattttaatgctcagctggtgcatttgggtgctatcaaggt	651		
Db 563	AAAN	622		
Qy 652	ttgacacgatgattactgatataatagcgaagtgcaaatgcacagtagtattgatatt	711		
Db 623	TKGGGTGTTTWTATKGTGTGGGKWTTKGKGKADTTACKTKGTGTTTTTTTTTATA	682		
Qy 712	tgtgtaagaagcaagagactaaacgcttcttaacagcagtaagtttaataaacagagatg	771		
Db 693	TKGGGTGGGGGRTKGGGRRATKGGGKGGTGTATGTGAKGGTGGTGTTRAGKTWAATG	742		
RESULT	5			
CNS00EVL	1101 bp	DNA	GSS	04-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence			
DEFINITION	BAC29B23 of RPI-98 library from drosophila melanogaster (fruit fly); genomic survey sequence.			
ACCESSION	AL069706			
VERSION	AL069706.1			
KEYWORDS	GSS.			
SOURCE	fruit fly			
ORGANISM	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 1101)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage			
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
	- web : www.genoscope.cns.fr)			
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pletter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,			

BACR08K08	of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL064091	GSS.
AL064091.1	GI:4941847
GSS.	
fruit fly,	Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
1 (bases 1 to 1101)	Genoscope.
Direct Submission	
Title	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr ; web : www.genoscope.cns.fr)
JOURNAL	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	Location/Qualifiers
source	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPCI-98"
	/clone="BACR08K08"
	/note="end : TET3"
BASE COUNT	395 a 120 c 103 g 334 t 149 others
ORIGIN	
Query Match	5.8%; Score 57; DB 219; Length 1101;
Best Local Similarity	38.1%; Pred.No. 0.011;
Matches 177; Conservative	50; Mismatches 233; Indels 4; Gaps 1;
Qy	165 tattgtaaagcgggatatcatcaatgcctttaaatttttagaatatttcacaaaagtgtcacaaagggtt 224
Db	605 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTAATTTTWTATNITTTTAAATTTTAAAAATAAATTTT 664
Qy	225 ttattacagcgccgatagaactagacaagtgtttttaccataacctatcgactgaattaataa 284
Db	665 TATTTAWNAAWAATTAATAAAAAAAAAAAAAAAAAAAAATATATAAAAAAAAAAATATATAAAATWT 724
Qy	285 gttaattcaatttgatgaaaaagcagatttaattattttaaacgcttttcctaacc 344
Db	725 AAATATATATTTTWAATAATATATTAATAAANAATAATTTTATTATATATAAATAATW 784
Qy	345 agtatgcggtactaactgagcaatttaacataattcccagttgcgcagtgatgac 404
Db	785 ATWATTATATATNNWNATAATTTTWTAAATTTTWATAATTAATTTMATTAATAAAATTTT 844
Qy	405 agactatcgcttacataaaaactggattacgcgctattccaagaatattatgtggcaac 464
Db	845 AWATATAAANAAWAAAAAATAATATAAATAATATWTATTTWTATAWTTTWWAAATAAATAAAT 904
Qy	465 aaagaacagcaagaagcttcatacgatgaggttaggttagcttccttcaacagcttaaatgac 524
Db	905 AAWAAAAAATAATAWATAAATAAATAAATAAATAATTTTWTATTAWAAAYWATTAWAWAWA 964
Qy	525 aggtattccctattgata----acaaattggaacgccctatttaatcaaaagcagtggttaa 580
Db	965 ATATTTTATATATWATTTATATATAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1024

Qy	308	aaagccgagatctaattattacagtttctoctacaccagtgtatcggtactaacgtgcgc	367
Db	839	AAATAAATTTTAAAAAWATTTAAGAAATTTWAATATWTAATAAAAAAAAAAATTAATAAAWAAWT	898
Qy	368	aa--ttaacattaataattccagtttgtctacagtgatcacagacctatcgcttacataaaaa	425
Db	899	WWTTTTTTTTTWAAWTTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAWAAAAAAA	958
Qy	426	ctggattacgcgcgtattcaacaagatatattgtggcaacaaaagaacgaacacagactt	485
Db	959	AATATAAANNAANNAWTTTAATTTTATTTTAAAAAANAANAATAAATAATAAATAAAT	1018
Qy	486	catagacgtaggattgatcctccaacagtttaagtcagcaggttatcctattgatacaaa	545
Db	1019	AAAAATTTTATATATTAATAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAA	1078
Qy	546	atttgaaa 553	
Db	1079	AAAAAAA 1086	
RESULT	10		
CNS016LI/c			
LOCUS	CNS016LI	1101 bp	DNA GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL106896		
VERSION	ALI06896.1	GI:5624374	
KEYWORDS	GSS:		
SOURCE	fruit fly.		
ORGANISM	Plasmid Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submission		
TITLE	Submitted (23-JUL-1999); Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
JOURNAL	- web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster PAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.		
FEATURES	Location/Qualifiers		
source	1..1101		
	/organism="Drosophila melanogaster"		
	/plasmid="pBelOBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN16D22"		
	/note="end : T7"		
BASE COUNT	203 a	220 c	84 g 158 t 436 others
ORIGIN			
Query Match	5.7%	Score 55.6;	DB 219; Length 1101;
Best Local Similarity	31.6%;	Pred. No. 0.022;	
Matches 101;	Conservative	Mismatches 144;	Indels 0; Gaps 0;
Qy	23	tattgatttaactgcctcattcgtagcgttcattgcacagtcattgcacagtcattgcgtta	82
Db	1101	TATTATATATATHYAYWYVYYTMYAWWWVTCTMTAMTAAANNAANNAANNAANNA	1042
Qy	83	atcaacttatgatgaatcagaccatttaagcgttcattgcacagtcattgcgttaattattcg	142
Db	1041	AWAWATTTATTTTHWTAAHTATWTWCAATAHHWAWWTTTATATWAATTTATTTATTAH	982

ACCESSION AL175696
VERSION AL175696.1 GI:7813753
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.

Search completed: June 29, 2001, 10:55:58
Job time: 5021 sec

REFERENCE 1 (bases 1 to 1092)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1092)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 1092)
Genoscope.

REFERENCE Direct Submission
AUTHORS Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..1092
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="222L11"
/clone_lib="G"
/note="Genoscope sequence ID : COAG222CF061L1-end : T7"

BASE COUNT 383 a 169 c 165 g 262 t 113 others
ORIGIN

Query Match 5.6%; Score 54.8; DB 220; Length 1092;
Best Local Similarity 33.7%; Pred. NO. 0.033;
Matches 128; Conservative 64; Mismatches 188; Indels 0; Gaps 0;

QY 5 ttactcaaaataaagatattgattattactgctcattcggttaacggtcatatgcaag 64
Db 992 TTTTAAAAATAAATAAATTTTNTTNTTATATTTTTTTTTTTTWTAAWAAAAWAAA 933

QY 65 ttacacagatcggttaactaatgatgatgaatcagaccatttaagcgcattg 124
Db 932 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 873

QY 125 agcagatttatttggaagctcccaatttgactctctatttgtaaaaaatggtata 184
Db 872 ATTWAAATTTTAAATTTTAAWAAATTTTAAWAAATTTTAAWAAATTTTAAWAAATTTT 813

QY 185 tcaatagcttaataatttagaataatgtacaaaggggtttattacagccgcagata 244
Db 812 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 753

QY 245 aactagacaaatgttttacaaactatgaccttaataagtttaattatttgataa 304
Db 752 AWTWTTTWT 693

QY 305 aagaaagccagatttaataattactgcttccacacagttatgctggtactaactg 364
Db 692 AAAAAAATTTTATTTTAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAA 633

QY 365 agcaatttaacattatatt 384
Db 632 TTWTTAAATAAATAAATAA 613

7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 11:46:43 ; Search time 119.12 Seconds
(without alignments)
1516.360 Million cell updates/sec

Title: US-09-668-788-3

Perfect score: 975

Sequence: 1 atggttactcaataataaaaa.....gtaaaacgtgatactccag 975

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	437	44.8	2600	3	US-08-988-251-1
C 2	437	44.8	2600	4	US-09-386-048-1
C 3	46.4	4.8	8920	2	US-08-446-855A-1
C 4	46.4	4.8	8920	4	US-09-150-741-1
C 5	44.8	4.6	7218	1	US-08-232-463-14
C 6	43.8	4.5	19124	2	US-08-487-826B-13
C 7	42	4.3	19124	2	US-08-487-826B-13
C 8	41	4.2	2058	2	US-08-749-391-1
C 9	41	4.2	2058	4	US-09-390-200-1
C 10	40	4.1	8654	1	US-08-920-812-6
C 11	40	4.1	8654	1	US-08-920-821-6
C 12	40	4.1	8654	1	US-08-921-177-6
C 13	40	4.1	8654	1	US-08-362-577C-6
C 14	40	4.1	8654	2	US-08-920-828-6
C 15	39.6	4.1	2022	2	US-08-505-486-96
C 16	39.6	4.1	2022	3	US-08-801-028-96
C 17	39.6	4.1	2022	3	US-09-340-154-96
C 18	39.6	4.1	2022	5	PCT-US95-09338-96
C 19	39.6	4.1	2022	5	PCT-US95-09338-96
C 20	39.6	4.1	2127	3	US-08-505-486-95
C 21	39.6	4.1	2127	3	US-08-801-028-95
C 22	39.6	4.1	2127	3	US-09-340-154-95
C 23	39.6	4.1	2127	5	PCT-US95-09338-95
C 24	39.6	4.1	2127	5	PCT-US95-09338-95
C 25	39.4	4.0	2960	3	US-08-913-842-3
C 26	39.4	4.0	8700	2	US-08-392-625-16
C 27	39.4	4.0	8700	2	US-08-466-961A-16

Sequence 18, Appl
Sequence 595, App
Sequence 972, App
Sequence 20, Appl
Sequence 21, Appl
Sequence 22, Appl
Sequence 36, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 193, App
Sequence 1014, Ap
Sequence 920, App
Sequence 8, Appl
Sequence 8, Appl
Sequence 1137, Ap

28 39.4 4.0 8700 2 US-08-645-193B-18
C 29 39.2 4.0 658 4 US-08-998-416-595
C 30 39 4.0 688 4 US-08-998-416-972
C 31 38.6 4.0 246240 2 US-08-724-394A-20
C 32 38.6 4.0 246240 2 US-08-724-394A-21
C 33 38.6 4.0 246240 2 US-08-724-394A-22
C 34 38.4 3.9 665 2 US-08-883-795A-36
C 35 38.4 3.9 2435 4 US-09-306-593-1
C 36 38 3.9 2950 5 PCT-US93-08386-7
C 37 38 3.9 2968 5 PCT-US93-08386-1
C 38 37.8 3.9 899 1 US-08-353-341-4
C 39 37.6 3.9 658 4 US-08-998-416-193
C 40 37.6 3.9 676 4 US-08-998-416-1014
C 41 37.6 3.9 677 4 US-08-998-416-920
C 42 37.6 3.9 1511 1 US-07-991-867B-8
C 43 37.6 3.9 1511 1 US-08-107-755A-8
C 44 37.6 3.9 1511 2 US-08-544-332-8
C 45 37.4 3.8 636 4 US-08-998-416-1137

ALIGNMENTS

RESULT 1
US-08-988-251-1/c
; Sequence 1, Application US/08988251
; Patent No. 6013497
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: Mure
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,251
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10094
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-988-251-1

Query Match 44.8%; Score 437; DB 3; Length 2600;
Best Local Similarity 100.0%; Pred. No. 5.5e-97;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttactcaataataaaagatatgtattacttggtcattcggtacggtcatatg 60
|||||

Db	8562	TTACGTTTAAATATTTAAATAAATGTTTTTATTAAATATGATCATTAATTTATTATGATTTA	8621
Qy	189	tgcgtttaaatattttgagaataattgacaaagggtttttattcacagcgcccgagataaac	248
Db	8622	TTTTTTTTATAAAATTTTTTGTTATATATACAAATTTTTATTATTCTACTCATATGTATAAAC	8681
Qy	249	agacaaatgtttttcacaaatactatggacttaataagtttaatttatttgata--aaa	306
Db	8682	AAAATGGTTTTTCAATTTACAAATAAATTTTATAATTTTAAATAATTTATTAAATATAA	8741
Qy	307	gaaagccagatttaataattatttaacgtttccctacacaggttatgtcgggtactaaactgag	366
Db	8742	AAAAATAAATATATATAAACATATAAAATGTATAAATCTTTTTTAATTATATAAATAATTAT	8801
Qy	367	caatttaacacttaatt	384
Db	8802	AAATGTTATGATTTTTTT	8819

```

4
RESULT
US-09-150-741-1
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09-150-741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/4446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

[illegible]

Db 8802 aaatgttatgatttttt 8819

RESULT 5

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEFFELINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Su

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,

FILING DATE:

APPLICATION NUMBER: Ep 91 114

FILING DATE: 126-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTgpt-F1s

US-08-232-463-14

[illegible]

[illegible]

RESULT 6
 US-08-487-826B-13
 ; Sequence 13, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Slim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellens, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelisen, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19124 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 US-08-487-826B-13

[illegible]

RESULT 7
 US-08-487-826B-13/c
 ; Sequence 13, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellens, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knodbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelsen, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19124 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHEetical: NO
 ; ANTI-SENSE: NO
 US-08-487-826B-13

	Query Match	4.3%	Score 42;	DB 2;	Length 19124;
	Best Local Similarity	47.1%;	Pred. No.	0.2;	
	Matches 129;	Conservative 0;	Mismatches 145;	Indels 0;	Gaps 0;
QY	132	t t t a t t t g g a a g c t a c c a a t t t g a c t t o r a t t t g t a a a a a a g g t a t a c c a a t a g	191		
DB	7095	T A T A T A T A T A T A T A T A T A T A G T T G A C T A T A T A T A T A T T A C A A N T A T G T A I A T	7036		
QY	192	c t t t a a a t a t t t t a g a a a t a t g t a c a a a g g g t t t a t t a c a g c c g c c a g a t a a c t a g a	251		


```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neocallimastix patriciarum
; STRAIN: 27
; IMMEDIATE SOURCE:
; LIBRARY: genomic DNA library
; CLONE: pNspX-06
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 301..1755
;
US-09-390-200-1

Query Match          4.2%; Score 41; DB 4; Length 2058;
Best Local Similarity 46.7%; Pred. No. 0.21;
Matches 163; Conservative 0; Mismatches 185; Indels 1; Gaps 1;

QY 1 atggttactcaaaataaagatatgattattacttggtcattcggttaacggtcatatg 60
   || || || || || || || || || || || || || || || || || || || ||
Db 9 ATAATTGTTCAAAAAGAGTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 68
   || || || || || || || || || || || || || || || || || || || ||

QY 61 caagttacacagagatcgtttaacacttaatatgatgaatcagaccatttaacgcgtc 120
   || || || || || || || || || || || || || || || || || || || ||
Db 69 TGAGTATAAATAGTTTGGTTTACCTTTTGGTTTTCCTTTATTCCTTATAAAGTTA 128
   || || || || || || || || || || || || || || || || || || || ||

QY 121 attgagcacgatttattatggagctcaccatttgactctctatttgtaaaaaatgg 180
   || || || || || || || || || || || || || || || || || || || ||
Db 129 ATTGTTTAATAAATATTGGTGGAAATATTAAAGTTGTATATATATATTAAATTTATT 188
   || || || || || || || || || || || || || || || || || || || ||

QY 181 tatatacagctttaaatatttagaataatgtacaaaggggtttat-tacagcgccc 239
   || || || || || || || || || || || || || || || || || || || ||
Db 189 GGAATTAATTACTTCTCGTGGACAAATAATTAATAGTGATATATATATTATTAGAA 248
   || || || || || || || || || || || || || || || || || || || ||

QY 240 agataaactagacaaatgtttttacaaataactatggacttaataagtttaattatt 299
   || || || || || || || || || || || || || || || || || || || ||
Db 249 AAAGAAAAAATAAATATTACAAATTAATTACTATAAATAAATAAATAAATAAATAA 308
   || || || || || || || || || || || || || || || || || || || ||

QY 300 gataaaagaaagccagatttaattatttaacggtttccctacacacagttt 348
   || || || || || || || || || || || || || || || || || || || ||
Db 309 TTTACAAATTAATTCCTGTATTATTATATCTTTAACTTCAACTACTCTTGCT 357
   || || || || || || || || || || || || || || || || || || || ||

RESULT 10
US-08-920-812-6/c
; Sequence 6, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus epidermidis
; STRAIN: Clinical Isolate SE-22
;
US-08-920-812-6

Query Match          4.1%; Score 40; DB 1; Length 8654;
Best Local Similarity 52.1%; Pred. No. 0.51;
Matches 88; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 195 taaatttttgaataatgtacaaaggggtttattacagccgccagataaactagacaa 254
   || || || || || || || || || || || || || || || || || || || ||
Db 1648 TAAAAAAATCGATGATTTACTTTAAAGAGTAACCAACCCGTATCACAGAAGCAATAATGA 1589
   || || || || || || || || || || || || || || || || || || || ||

QY 255 atgttttttcaaaactactggacttaataagtttaattattattgataaaagaaagcc 314
   || || || || || || || || || || || || || || || || || || || ||
Db 1588 AATGAAGTATAATGCTAGGAGTNAACACAGATATTTCTTTCACCTTATGAAAGATAAGAC 1529
   || || || || || || || || || || || || || || || || || || || ||

QY 315 agatttaattatttaacggtttccctacaccagttatgctggtactaaact 363
   || || || || || || || || || || || || || || || || || || || ||
Db 1528 TGTTTTACTCTTTTATTAAAGTTAATACGATAGTATCTGAGTCCAAACT 1480
   || || || || || || || || || || || || || || || || || || || ||

RESULT 11
US-08-920-827-6/c
; Sequence 6, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
```

REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8654 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORGANISM: Staphylococcus epidermidis
STRAIN: Clinical Isolate SE-22
US-08-920-827-6

Query Match 4.1%; Score 40; DB 1; Length 8654;
Best Local Similarity 52.1%; Pred. No. 0.51;
Matches 88; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 195 taaatattttgagaatgtacaaagggtttttattacagcgcccgataaaactagacaa 254
DB 1648 TAAAAAATCGATGATTTACTTTAAACGAGTAACACCGGTATCACAGAGCAAAATATGA 1589
QY 255 agttttttacaaatacttgacttaataagtttaattatttgataaaagaaagcc 314
DB 1588 AATGAAGTATATGCTAGGGAGTAAACAGATATTTCTTCACATTATGAAGATAAGAC 1529
QY 315 agatttaataattattacagttctctacaccagttatgtcggtaact 363
DB 1528 TGTTTTACTCTTTTAAAGTTAATACGATAGTGTGAGTCCAACT 1480

RESULT 12
US-08-921-177-6/c
Sequence 6, Application US/08921177
Patent No. 5798211
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448

TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8654 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORGANISM: Staphylococcus epidermidis
STRAIN: Clinical Isolate SE-22
US-08-921-177-6

Query Match 4.1%; Score 40; DB 1; Length 8654;
Best Local Similarity 52.1%; Pred. No. 0.51;
Matches 88; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 195 taaatattttgagaatgtacaaagggtttttattacagcgcccgataaaactagacaa 254
DB 1648 TAAAAAATCGATGATTTACTTTAAACGAGTAACACCGGTATCACAGAGCAAAATATGA 1589
QY 255 agttttttacaaatacttgacttaataagtttaattatttgataaaagaaagcc 314
DB 1588 AATGAAGTATATGCTAGGGAGTAAACAGATATTTCTTCACATTATGAAGATAAGAC 1529
QY 315 agatttaataattattacagttctctacaccagttatgtcggtaact 363
DB 1528 TGTTTTACTCTTTTAAAGTTAATACGATAGTGTGAGTCCAACT 1480

RESULT 13
US-08-362-577C-6/c
Sequence 6, Application US/08362577C
Patent No. 580673
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8654 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA

Query Match	4.1%	Score 39.6;	DB 2;	Length 2022;
Best Local Similarity	45.3%;	Pred. No. 0.45;		
Matches 167;	Conservative 0;	Mismatches 201;	Indels 1;	Gaps 1;
23	tattgattattactggtccttcggtgaacgctcatatgcaggttacacagagtagtcgta	82		
386	TATCAAAATATAATTTTGTAGTGACGTGTGTATAATTTAAAAAAAATTCACCTCTTCGTTTATAA	327		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:59:02 ; Search time 35.83 Seconds

(without alignments)
661.568 Million cell updates/sec

Title: US-09-668-788-4

Perfect score: 2048

Sequence: 1 MVTQNKILLINGSFGNHM.....SSQPOEIVGKVPYARFFVK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	591.5	28.9	382	AA133444	B. subtilis glycosyl transferase catalytic domain.
2	296.5	14.5	468	AA19042	Amino acid sequence
3	284	13.9	533	AA19041	Amino acid sequence
4	283	13.8	422	AA19040	Monogalactosylidase
5	283	13.8	525	AA19040	Amino acid sequence
6	279.5	13.6	522	AA19039	Amino acid sequence
7	272	13.3	492	AA19039	Arabidopsis thaliana
8	272	13.3	551	AA19039	Arabidopsis thaliana
9	252	12.3	404	AA19039	Arabidopsis thaliana
10	139	6.8	374	AA19039	E. coli colitose o
11	119	5.8	352	AA19039	Streptococcus pneumoniae

12	109.5	5.3	534	13	AA19039	HUG-Br2. Homo sapiens
13	106	5.2	530	21	AA19039	Human UDP-glucuron
14	102.5	5.0	934	16	AA19039	Human MSH2 protein
15	102.5	5.0	934	16	AA19039	Human MSH2. Homo
16	102.5	5.0	934	18	AA19039	Mismatch repair pr
17	102	5.0	530	19	AA19039	Uridine diphospho-
18	101.5	5.0	367	21	AA19039	Streptococcus pneu
19	98.5	4.8	2366	17	AA19039	C. difficile toxin
20	98.5	4.8	2366	19	AA19039	Clostridium diffi
21	98	4.8	510	13	AA19039	Japanese Black Pin
22	96.5	4.7	533	13	AA19039	HUG-Br1. Homo sap
23	96	4.7	662	18	AA19039	H. pylori cytoplas
24	95.5	4.7	1435	20	AA19039	PolC gene product
25	95.5	4.7	1435	22	AA19039	Amino acid sequenc
26	95	4.6	467	21	AA19039	Arabidopsis thalia
27	94.5	4.6	484	21	AA19039	Arabidopsis thalia
28	94.5	4.6	484	21	AA19039	Arabidopsis thalia
29	92.5	4.5	427	21	AA19039	Arabidopsis thalia
30	92.5	4.5	1151	20	AA19039	Human N-arginine d
31	92	4.5	648	19	AA19039	Helicobacter methi
32	92	4.5	1829	18	AA19039	DNA polymerase wit
33	90.5	4.4	503	21	AA19039	Streptococcus pneu
34	90	4.4	1802	21	AA19039	Plasmodium falci
35	89	4.3	350	19	AA19039	Streptococcus pneu
36	89	4.3	1049	21	AA19039	Human ORF1889
37	89	4.3	1092	19	AA19039	Staphylococcus epi
38	89	4.3	1174	21	AA19039	Lipid metabolism p
39	88.5	4.3	366	21	AA19039	Arabidopsis thalia
40	88	4.3	487	19	AA19039	H. pylori GHPO 479
41	88	4.3	666	18	AA19039	Herbicide-resistan
42	87.5	4.3	487	22	AA19039	Exophiala spinifer
43	87.5	4.3	873	21	AA19039	Arabidopsis thalia
44	87.5	4.3	2120	21	AA19039	Streptococcus pneu
45	87	4.2	397	21	AA19039	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AA133444
ID AA133444 standard; Protein: 382 AA.

XX

XX AA133444;

XX AC

DT 13-DEC-1999 (first entry)

XX

XX B. subtilis glycosyl transferase catalytic domain.

DE

XX Catalytic domain; glycosyl transferase; processive activity; detergent;

KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;

KW food emulsifier; polymer modifier; glycosyl transferase.

XX

XX Bacillus subtilis.

XX

PN DE19819958-Al.

XX

XX 30-SEP-1999.

XX

PF 05-MAY-1998; 98DE-1019958.

XX

XX 25-MAR-1998; 98DE-1013017.

XX

XX (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.

PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.

XX

PI Wolter FP, Jorasch P, Heinz E, Zaehlinger U;

XX

DR WPI; 1999-552364/47.

XX N-PSDB; AA223386.

XX

PT New glycosyl transferase protein, useful for producing glycolipids

XX

PS Example 1; Page 7; 8pp; German.

XX This invention describes a novel protein (I) comprising identical or
CC different catalytic domains of glycosyl transferases and which has
CC processive activity (i.e. builds up oligosaccharide glycolipids by
CC successive addition of glucose units). (I) can be used to produce
CC 3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
CC modifiers or detergents. This sequence represents a Bacillus subtilis
CC glycosyl transferase catalytic domain which is used in the method of
CC the invention.

XX Sequence 382 AA;

Query Match 28.9%; Score 591.5; DB 20; Length 382;
Best Local Similarity 35.7%; Pred. No. 8.4e-48;
Matches 136; Conservative 77; Mismatches 131; Indels 37; Gaps 10;

QY 5 NKKILITGSGFNGHMQVTQSVINQNDMLDHSVIEHDLFMEAHPIILTSICKKWIINS 64

Db 4 nkrvllltanygnghvqvaktlyeqcgrlqfghvtv--snlygesnplvsevtqlylks 61

QY 65 FKFRNMVKGFFYSRPDKLDKCFYK-----YYGL-NKLINLLIKE-KPDLILLTFPTPM 117

Db 62 fsigkqfyrlyfyg-----vdklynrkrkfnlyfkmgnkrigelvdehqpdliintfpmivv 117

QY 118 SVLTQFNINIPVATVMDYRLHKNWITPYSTRYVATKTKQDFIDVGDIDPSTVKVTGI 177

Db 118 peyrrtgrvptfvmtdfclhkwthenvdkyyvatdyvkekilleigthpsnvkitgi 177

QY 178 PIDNKFT-----PINQKOWLIDNNLDPDKQITILMSAGAFVSGKFDFTMITDILAKSANA 232

Db 178 pirpqqfeesmpvpykky-----nlspnkvvllimagahgvlknvkelcen-lvkddqv 231

QY 233 QVVMICGKSKELKRSITA-----KFKLTRMYLILGYTKHNMENWASQOLMITKPGGIT 285

Db 232 qvvvcvgktaikeslsaleaengdkl-----vlgvyveridelfrtdcmthkpggit 285

QY 286 ITEGFARCIPIFMFLNAPGQLENAFYFEKGFKGIAIDTPEEAIKIVASLTNGNQLTNM 345

Db 286 lteataigpvllykvpqgekenanffedrgaaivvnvrheilesvtlladedtlhrm 345

QY 346 ISTMEODKIKYATQICRDL 366

Db 346 kknikdlhianssevilleidil 366

RESULT 2

AAB19042

ID AAB19042 standard; protein; 468 AA.

XX

AC AAB19042;

XX 08-FEB-2001 (first entry)

DE Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.

XX Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;

KW apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;

KW acquired immune deficiency syndrome; coccidiosis.

XX Arabidopsis thaliana.

XX WO200056919-A1.

XX 28-SEP-2000.

XX 17-MAR-2000; 2000WO-FR00658.

XX 19-MAR-1999; 99FR-0003434.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PI Marechal E, Block M, Joyard J, Douce R;

XX WPI; 2000-602227/57.

XX Use of monogalactosyl diacylglycerol synthase for identifying its
PT specific inhibitors, potentially useful as antiparasitic agents and
PT herbicides

PS Example 2; Fig 2; 33pp; French.

XX The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC synthase. MGDG is present in all plastids tested and is essential for
CC cell survival, but is not present in other membrane structures, or in
CC animal cells, and so represents a specific target. MGDG synthase or a
CC plastid membrane isolated from plants is used to select and screen for
CC specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC active agents against apicomplex parasites or as herbicides. The
CC inhibitors are used as antiparasitic agents, especially against
CC Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC as herbicides.

XX Sequence 468 AA;

Query Match 14.5%; Score 296.5; DB 21; Length 468;
Best Local Similarity 25.4%; Pred. No. 1.e-19;
Matches 100; Conservative 78; Mismatches 173; Indels 43; Gaps 14;

QY 4 QNKKILITGSGFNGHMQVTQSVINQNDMLDHSVIEHDLFMEAHPIILTSICKKWIYN 63

Db 67 rtknvlilmsdtggghrasaeairdafklefgdkyrvivkdwke-----ytwgpln 118

QY 64 ---SFKXF---RNMVK-CFYVSRPDKLDKCFY---KYVGLNKLINLLIKEKPDILLT 111

Db 119 dmersykmvkhvqlwkvafstspkwhscylaalaayya-keveaglmeykpeilisv 177

QY 112 FP---TPVMSVLTQFNINIPVATVMDYR-LHKNWITPYSTRYVATKTKQDFIDVG 166

Db 178 hplmqhlpilwllkwelqkrlvltvtdlntchptwfpvncrcycpsqevakralfdg 237

QY 167 IDPSVVKVTGIPIDNKFTPINQKOWL-IDNNLDPDKQITILMSAGAFV-----SKFSD 219

Db 238 ldesqrvfvgiprpsfaravlkddlrkelemdqdlravllmggggmppvketakale 297

QY 220 TMITDILAKSANAQVVMICGKSKELKRSITA-KFKLTRMYLILGYTKHNMENWASQOLMI 278

Db 298 efilydkenrkpiqgmvcgrnkklaaleaidwkip--vkurgfetqmekwmgacdcil 355

QY 279 TKPGGITITEGFARCIPIFMFLNAPGQLENAFYFEKGFKGIAIDTPEEAIKIV----AS 334

Db 356 tkagpgtiaeslirslpilndyipgqekgnvpyvengagvftsrpsketarivgewfst 415

QY 335 LTNGNEQLTNMISTMEODKIKYATQICRDL 368

Db 416 ktdeleqtsdnarklaqpe---avfdlvkddidel 446

RESULT 3

AAB19041

ID AAB19041 standard; protein; 533 AA.

XX

AC AAB19041;

XX 08-FEB-2001 (first entry)

DE Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.

XX Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;

KW apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;

KW acquired immune deficiency syndrome; coccidiosis.

XX Arabidopsis thaliana.

XX PN W0200056919-A1.
 XX PD 28-SEP-2000.
 XX PF 17-MAR-2000; 2000WO-FR00658.
 XX PR 19-MAR-1999; 99FR-0003434.
 XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX PI Marechal E, Block M, Joyard J, Douce R;
 XX DR WPI; 2000-602227/57.
 XX PF Use of monogalactosyl diacylglycerol synthase for identifying its
 PT specific inhibitors, potentially useful as antiparasitic agents and
 PT herbicides -
 XX PS
 XX PA Example 2; Fig 2; 33pp; French.
 XX CC The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
 CC synthase. MGDG is present in all plastids tested and is essential for
 CC cell survival, but is not present in other membrane structures, or in
 CC animal cells, and so represents a specific target. MGDG synthase or a
 CC plastid membrane isolated from plants is used to select and screen for
 CC specific inhibitors of MGDG synthase. These inhibitors are suitable as
 CC active agents against apicomplex parasites or as herbicides. The
 CC inhibitors are used as antiparasitic agents, especially against
 CC Plasmodium, Toxoplasma (particularly in patients with acquired immune
 CC deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
 CC as herbicides.
 XX SQ Sequence 533 AA;

Query Match 13.98; Score 284; DB 21; Length 533;
 Best Local Similarity 25.5%; Pred. No. 2.1e-18;
 Matches 103; Conservative 66; Mismatches 185; Indels 50; Gaps 13;
 QY 6 KKIILITSGFNGHMVQTOSTIVNQLNDNLDHLSVIEHDLFMEAHPIILTSICKKWIYNSF 65
 Db 142 kvllmsdtggghrasaeairaafnqefgdeyqvftldlwdhtp-----wpfnql 193
 QY 66 KYFRN-----MYKGFYYSRPDKL----DKCFYKYGLNKLINLLIKEKPDILLTFP- 113
 Db 194 prsynflvkhgtlwkmtvygtsprlvhgsnfatstfiareiaaglmkyqpdililsvhpl 253
 QY 114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNMITYSTRYVATKTKQDFIDVGIDP 169
 Db 254 mqhvpilrvlrskgllkklvftvtldlstdtchptwfhklvtrcycpstevakraqaglet 313
 QY 170 STVKVTGIPIDNKFETPINOWL-----IDNLDLPDKQITILMSAGAGV-----SKGF 218
 Db 314 sqikvyglpvrpsvkprkvelrrelgmdenl----pavllmggggmgpieataral 369
 QY 219 DTMITDILAKSANAQVVMICKSKELSLTA-KFKLTRMYLILGYTKHNMENWASSQLM 277
 Db 370 adalydknlgveavqgvllicgrnklslssldwkip--vqvkgfittkmeecmgacdci 427
 QY 278 ITKPGGITITEGFARCIPIFMFLNPAPGOELENAFYFEKGFKIADTPPEAKIVASLTN 337
 Db 428 itkagpgtiaeamirgipilingyaggeagnvpyvengcgkfskpskelisvadwfg 487
 QY 338 GNEQLTNMISTMEQDKIYATQ-----TICRDLLDLI-GHSSQPP 376
 Db 488 ---paskelmsqnalrlakpeavfkivhdmhelvrkksnlpq 528

RESULT 4
 AAW41148
 ID AAW41148 standard; Protein; 422 AA.
 XX

AC AAW41148;
 XX 05-MAY-1998 (first entry)
 XX Monogalactosyldiacylglycerol synthetase sequence.
 XX Monogalactosyldiacylglycerol synthetase; MGDG; lipid production;
 XX cucumber.
 XX Cucumis sativus.
 XX JF10014579-A.
 XX 20-JAN-1998.
 XX 02-JUL-1996; 96JP-0172337.
 XX 02-JUL-1996; 96JP-0172337.
 XX (KIRI) KIRIN BREWERY KK.
 XX WPI; 1998-138241/13.
 XX N-PSDB; AAV12734.
 XX New mono:galactosyl-di:acyl-glycerol synthase - useful for, e.g.
 PT increasing lipid production in transformed host
 XX Claim 1; Page 8; 13pp; Japanese.
 XX This sequence represents the monogalactosyldiacylglycerol synthase (MGDG)
 CC of the invention. This sequence was isolated from cucumber. MGDG may be
 CC used to increase lipid production in an organism.
 XX SQ Sequence 422 AA;

Query Match 13.88; Score 283; DB 19; Length 422;
 Best Local Similarity 25.3%; Pred. No. 1.9e-18;
 Matches 105; Conservative 65; Mismatches 185; Indels 60; Gaps 13;
 QY 6 KKIILITSGFNGHMVQTOSTIVNQLNDNLDHLSVIEHDLFMEAHPIILTSICKKWIYNSF 65
 Db 33 kvllmsdtggghrasaeairaafnqefgdeyqvftldlwdhtp-----wpfnql 84
 QY 66 KYFRN-----MYKGFYYSRPDKL----DKCFYKYGLNKLINLLIKEKPDILLTFP- 113
 Db 85 prsynflvkhgtlwkmtvygtsprlvhgsnfatstfiareiaaglmkyqpdililsvhpl 144
 QY 114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNMITYSTRYVATKTKQDFIDVGIDP 169
 Db 145 mqhvpilrvlrskgllkklvftvtldlstdtchptwfhklvtrcycpstevakraqaglet 204
 QY 170 STVKVTGIPIDNKFETPINOWL-----IDNLDLPDKQITILMSAGAGV-----SKGF 218
 Db 205 sklkvlgpvrpsvkprkvelrrelgmdenl----pavllmggggmgpieataral 260
 QY 219 DTMITDILAKSANAQVVMICKSKELK---RSLTAKFKLTRMYLILGYTKHNMENWASSQ 275
 Db 261 skalydenhgepvgvllicgrnklslssldwkip--vqvkgfittkmeecmgacd 316
 QY 276 LMITKPGGITITEGFARCIPIFMFLNPAPGOELENAFYFEKGFKIADTPPEAKIVASLTN 335
 Db 317 ciltkagpgtiaeamirgipilingyaggeagnvpyvengcgkfskpskelisvadwfg 376
 QY 336 TNGNEQLTNMISTMEQDKIYATQ-----TICRDLLDLI-GHSSQPPQEIYGVKVPYLA 386
 Db 377 fgpk---adelimsqnalrlarpeavfkivhdmhelvrkksnlpq-----fvpqys 421
 RESULT 5
 AAW41148
 ID AAW41148 standard; protein; 525 AA.
 XX

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149502.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.

Query Match 13.3%; Score 272; DB 21; Length 492;
Best Local Similarity 24.9%; Pred. No. 2.6e-17;
Matches 104; Conservative 71; Mismatches 184; Indels 58; Gaps 16;

QY 6 KKLIITGSGNGCHMQVTSIVNQLN-DMNLDHLSVIEHDLFME-----AHPILTS 55
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 83 Kkvllmsdtggghrasaaairafndefgdyqsgle---filekigfrcvlhvftid 139
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 56 IC---KKWYINSFKYPRN-----MYKGYYSRPDKL----DKCFYKYKYGLNKLINLLI 101
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 140 lwtldhtpwpfnqlprsynflvkhgltlwkntyygtspriyhgnsfaatstfiarelaggml 199
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 102 KEKPDILLITFP---TPVMSVLTEQFINIPVATVMTDY-RLHKMWITPYSTRYYVATK 156
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 200 kygdhliisvhpqlmhqvprlvirskglklkvitvtldtschtptwfhkivtrcycpst 259
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 157 ETRQDFIDVGIDPSTVKVTGIPIDKNFETPINQKWL-----IDNNLPDQKTILMSAGA 211
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 260 evakraqakagletsqikvyglpvrrpsfvkpvrpkvelrrelgmdeul----pavilmggg 315
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 212 FGV-----SKGFDTMITDLAKSANAQVVMICGSKEKLKSLTA-KEKLTFRMYLILCYT 264
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 316 egmgiptearaladalydknlgeavgvqliicgrnknklsksldwkkip--vqvkqifi 373
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 265 KHMENWASSQLMTIKPGGITITEGFARCIPIFMFLNPAPGQELENAFYFEKGFGKIADT 324
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 374 tkmeecmgacdciiitkagpgtiaeamirglpilngiyaggaagnpyvvengcgkfks 433
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 325 PEEAIKIVASLTNGNQNLNMISTMEQDIKVATQ-----TICRDLDDII-GHSSOPQ 376
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 434 pkeisklivadvfg--paskelismsqnalrlakpeavfkivdhdmhelvrkknslpq 487
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

RESULT 8
AAG42413
ID AAG42413 standard; Protein; 551 AA.
XX AC AAG42413;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 52893.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.


```
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.3%; Score 272; DB 21; Length 551;
Best Local Similarity 24.9%; Pred. No. 3.le-17;
Matches 104; Conservative 71; Mismatches 184; Indels 58; Gaps 16;

QY 6 KKLIITGSGNGHMQVTQSIIVNQLN-DNNLDHLSVIEHDLFME-----AHPILTS 55
DB 142 kvllmsdtggghrasaeairaafnqfgeyqsgle---filelekigfrcvlhvftid 198
QY 56 IC---KKWYINSFKYFRN-----MYKGFYYSRDKL----DKCFYKYGYGLNKLINLLI 101
DB 199 lwtthtpwfnqlprsynflvkhgtlwkmttygtsprihvqsnfaatatstfiarelaqglm 258
QY 102 KEKPDILLITFP----TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPXYSTRYYVATK 156
DB 259 kyqpdililsvhplmqhvlprlrsrgllkivftvitdltstcptfhkivtrcycpst 318
QY 157 EFKQDFIDVDIPSVKVTGIPIDNKFETPINOKOWL-----IDNNLDPDKOTILMSAGA 211
DB 319 evakraqkagletsikygipvrpsfvkprkvelrrelgmdehl----pavllmggg 374
QY 212 FGV-----SKGDFPMITDILAKSANAQVMTGSKSKELKSLTA-KFKLFRMYLILGYT 264
DB 375 egmgpieataraladalydknlgeavgvllicgrnkkqlsklssldwkip--vqvkgfi 432
QY 265 KHMENWMASSQLMITKPGGITTEGFCARCPIMFLNPAPGOELENAFYFEKGGFKIADT 324
DB 433 tkmeecmgacdciitkpggtiaeamirgplliingyiaagdeagvnyvvengcgkfsks 492
QY 325 PEEAKIVASLTNGNEQLTNMISTMEQDKIKYATQ-----TICRDLDDLI-GHSSOPQ 376
DB 493 pkeiskiradvfg---paskelmsqalriakpeavfkivhdmhelvrkkslbpq 546

RESULT 9
ID AAG42415
XX AAG42415 standard; Protein; 404 AA.
AC AAG42415;
XX
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 52895.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
```

```
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR
```


QY 169 PSTVKVTGIPIDNKFTTPINQKWLID--NNLDPDKQTILMSAGAFGVSKGFDITMIDIL 226
Db 159 gaitkvs---dqknp---dldvdiqthfhkhiptvlfvgssag-arvfnqlvtdhk 209
QY 227 AK-SANAQVVMICGKS--KELKRSRTAKFKLTRMYLILGYTKHMEWMASSQLMITKPGG 283
Db 210 keiteryniinlkgdsslnelsqnlfrvdytdlyqpl-----meladvvtrrga 260
QY 284 ITTEGFA----RCIPMIFLNPAPGOELENAFYFEKFGKIADTPEEAIKIVASTNGN 339
Db 261 ntifellataklhviplgreasrgdienaayfvkkyae--dlqesdl-----tidsle 314
QY 340 EQITNMISTMEQDKIKYATOTICROLLD 367
Db 315 ekishllshkedyqakmkaskelksiad 342

RESULT 12
ID AAR26154 standard; Protein; 534 AA.
XX AAR26154;
AC AAR26154;
DT 27-JAN-1993 (first entry)
DE HUG-Br2.
XX
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide.
XX
OS Homo sapiens.
XX
FH Key
FT Region 12..22 Location/Qualifiers
FT /note= "putative membrane-insertion signal"
FT 492..508
FT /note= "putative membrane-anchoring peptide"
FT 348
FT Modified-site
FT /note= "predicted Asn-linked glycosylation site"
FT Misc-difference 282..285
FT /note= "residues encoded by TGCCACACGGGAAG !"
XX
PN W09212987-A.
XX
XX 06-AUG-1992.
XX
XX 10-JAN-1992; 92WO-US00282.
XX
XX 10-JAN-1991; 91US-0639453.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Owens IS, Ritter JK;
XX
XX WPI; 1992-284593/34.
XX N-PSDB; AAQ27369.
XX
XX Isolated gene locus UGT1, DNA segments and diagnostic probes -
XX for diagnosing Gilbert's disease and Crigler-Najjar syndrome
XX types I and II
XX
XX Disclosure; Fig 9A-I; 99pp; English.
XX
XX Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
XX been isolated. They are referred to as HUGBr1 (AAQ27369) and HUGBr2
XX (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
XX upon expression individually in COS-1 cells, encode isoforms that
XX catalyse the formation of the two bilirubin monoglucuronides and
XX the diglucuronide.
XX The cDNAs contain identical 3' ends (1469 bp in length) to each
XX other and to that of the human phenol transferase cDNA, HLUGP1
XX (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).

CC In contrast, they have unique 5' ends.
XX
SQ Sequence 534 AA;

Query Match 5.3%; Score 109.5; DB 13; Length 534;
Best Local Similarity 20.4%; Pred. No. 0.078;
Matches 58; Conservative 54; Mismatches 108; Indels 65; Gaps 12;
QY 93 LNKLINLIIKEKPDILLTPTPTVMSVLTEQFNINIPVATVMTDYLRLHKMWTPTSTRYY 152
Db 208 lgrvkmlylalsyichftsfasyaslasefgrvctvgdlssasv---wl----- 256
QY 153 VATKETKODFI---DVGIDPSTVKVTGI-----PIDNKEETPINQKWLIDNNLDPDKQ 203
Db 257 -----frsdrvkyprpimnmvfigincangkpisqefeyain----- 296
QY 204 TILMSAGAFG-VSKGFDITMIDILAKSANAQVVMICGSKELKRSRTAKFKLTRMYLILG 262
Db 297 ----asgehgiwvfvfslslesmvseipekkama----iadalgkipqvtvlwrygttrpsnlan 348
QY 263 YTKHMEWMASSQLM-----ITKPGGITITEGFARCIPIFMIFLNPAPGOELENAFYFE 314
Db 349 nt-ilvkwlpqndllghpmtrafithagshgvyescngvpmvmm-plfgdqmdnakrme 406
QY 315 EKGFGKIAD---TPBEAIVASLTNGNEQLTNM--ISTMEQDK 353
Db 407 tkgagvtlnvlemtsedlenaqkavindksykenimrlslhdkdr 451

RESULT 13
ID AAY78935 standard; Protein; 530 AA.
XX AAY78935;
AC AAY78935;
DT 05-JUN-2000 (first entry)
DE Human UDP-glucuronosyltransferase 2B15 amino acid sequence.
XX
KW UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism: metabolism;
KW drug interaction; detect; human; single nucleotide polymorphism; SNPs.
XX
XX Homo sapiens.
XX PN W0200006776-A1.
XX
XX PD 10-FEB-2000.
XX
XX PF 22-JUL-1999; 99WO-US16675.
XX
XX PR 28-JUL-1998; 98US-0094391.
XX
XX (AXYS-) AXYS PHARM INC.
XX
XX PI Galvin M, Miller A, Penny L, Riedy M;
XX
XX WPI; 2000-195321/17.
XX DR N-PSDB; AA295206.
XX
XX PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
XX genotyping individuals to predict rate of metabolism of substrates and
XX PT for identifying potential drug interactions -
XX
XX PS Disclosure; Page 59-60; 72pp; English.
XX
XX This sequence represents the human UDP-glucuronosyltransferase 2B15
XX (UGT2B15) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
XX family of enzymes that catalyse the glucuronic acid conjugation of a
XX wide range of endogenous and exogenous substrates. The UGT2B gene
XX subfamily encode steroid metabolizing isoforms in the liver. Alteration
XX of the expression or function of UGTs may effect drug metabolism. The
XX CC invention relates to non-chromosomal nucleic acid molecules, which

CC comprise human UGT2B sequence polymorphisms (see AA295051-295110). Probes
CC which detect the UGT2B locus polymorphisms can be used to detect altered
CC UGT2B metabolism of a substrate in an individual. The nucleic acid
CC molecules comprising a human UGT2B sequence polymorphism can be used in
CC screening assays for genotyping individuals, also to predict their rate
CC of metabolism of UGT2B substrate, potential drug-drug interactions and
CC adverse side effects. The polymorphisms can be used as single nucleotide
CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
CC variation in activity or expression of UGT2B protein. The polymorphism
CC containing nucleic acid molecules may also be used for generating
CC genetically modified non-human animals and for obtaining site specific
CC gene modification in cell lines.

XX Sequence 530 AA;

Query Match 5.2%; Score 106; DB 21; Length 530;
Best Local Similarity 19.2%; Pred. No. 0.16; Indels 154; Gaps 24;
Matches 93; Conservative 78; Mismatches 160; Indels 154; Gaps 24;

Qy 1 MVTQNKILITGSGNGHMQVTSIVNQLNDMLDHLVSIEHDLFEAHPIILTSICKKW 60
Db 46 lvqrghevtltss-----astlvnask-----saik-----levyp--tsltknd 85
Qy 61 YINSP-----KYFRNMVKGFFYSRDKDKCFYKYYG-----LNKLINLLIKE-K 104
Db 86 ledslilkildrwiygskntfswysfqlqelcweydydsklckdavlknklmmklgesk 145
Qy 105 PDLILLTPTPVMVSLTQFNI-----NIPVA-TVMTDY 137
Db 146 fdlvlladalncpgellaelfnlpflsrfsvgytfeknsggfflppsvvpvmseisdq 205
Qy 138 RLH-----KN-----WIPYSTRYVATKQKQDFIDVGPSTVVKVTGIPIDNKF 184
Db 206 miferiknmihmlydfwfygd-----lkkwdqfysevlgrpttl-----fe 249
Qy 195 TPINQKQWLIDN-----NLD-----PDQOTILMSAGAFGV---S 215
Db 250 tmgaemwllrtywdfefrplnvdfvgqlhckpakplpkemeefvsgsgengivvfv 309
Qy 216 KGFDPMITDILAKSANAQVWVICGSKELKRSLSLAKFKLTRMYLILGYTKHNMWMASSQ 275
Db 310 lg-smismseesan-----miasaladipkvilwrf-gkpnltlgsntlykwlpgnd 362
Qy 276 LM-----ITKPGGITTEGFARCIPIFNLNAPQGELENAFYFEKGFKIADTPEE 327
Db 363 llghpkkafithggtngiyaeiylgipmvgi-plfadqhdniakmkag-----411
Qy 328 AIKIVASLTNGEQLTNMISTMEQDKI-KYATQTCRDLDDLIGHSSQPOQEIYGVKPLYA 386
Db 412 aalsvdlrtmsrdllnalkavindpykenvmklisri-----hhdqpmkpldravfw 465
Qy 387 RFEVK 391
Db 466 efvmr 470

RESULT 14
AAR76063
ID AAR76063 standard; Protein; 934 AA.

XX AAR76063;

XX 15-JAN-1996 (first entry)

XX Human MSH2 protein, homologue of the E.coli mutS gene product.

XX Mismatch repair; MSH2; primer; identification; defect; alteration;
XX cancer; tumour; vaccine.

XX Homo sapiens.

XX W09514085-A2.

PN

XX 26-MAY-1995.
XX 17-NOV-1994; 94WO-US13385.
XX 13-JUN-1994; 94US-0259310.
XX 17-NOV-1993; 93US-0154792.
XX 07-DEC-1993; 93US-0163449.
XX (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
XX (DAND) DANA FARBEN CANCER INST.
XX Fishel R, Kolodner RD, Reenan RAG;
XX WPI; 1995-200377/26.
XX N-PSDB; AAQ93901.
XX Determining alteration in human mismatch repair pathways - used in
XX the diagnosis, prognosis and therapy of cancers and in screening
XX assays
XX Claim 27; Page 159-163; 256pp; English.
XX AAQ93901 is the human mismatch repair pathway gene MSH2. Defects or
XX alterations in such a gene result in the accumulation of unstable
XX repeated DNA sequences, a feature of a number of different cancers.
XX The identification of a defect in the mismatch repair pathway can
XX be diagnostic of a predisposition to cancer and prognostic for a
XX particular mammalian cancer e.g colorectal, ovarian, endometrial
XX (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
XX sequences and polypeptides of the hMSH2 gene may also be used for
XX therapy and in vaccines.
XX Sequence 934 AA;

Query Match 5.0%; Score 102.5; DB 16; Length 934;
Best Local Similarity 22.3%; Pred. No. 0.81;
Matches 77; Conservative 45; Mismatches 131; Indels 93; Gaps 17;

Qy 18 GHMVOVSIVNQLNDMLN--DHLVSIE-----HDLFEAHPIILTSICKKWYI 62
Db 338 ggrlvnqwikplndknrieerlnlveafvedaelrqlqedl-lrrfpdlorlakk--- 393
Qy 63 NSFYFRNMVKGFFYSRDKDKCFYKYYGKLNKLINLL-IKEKPD-----LILLTFPPPV 116
Db 394 -----fgrqaanlqdcyrlvgginqipnvigalekhegkhqkillavfvtp 440
Qy 117 MSLTE--QFNINIPVATVMTDYRLHKWNIPTSTRYVATKQKQDFIDVGPSTVKV 174
Db 441 tdlrsdfskfqgemiettdmdqvenheflvxp-----sfqpnlsel 481
Qy 175 TGI--PIDNKE-TPINQKQWLIDNLDNPDQOTILMSAGAFVSGKGFDTMITDILAKSAN 231
Db 482 reimndlekkmqstlisaar---dlgidpgkqikldssaagfyfrvtckeekvlrnkn 538
Qy 232 AQVVMICGSKELKRSLSLAKFKLTRMYLILGYTKHNMWMASSQLMITKPGGITTEGFA 291
Db 539 fstvdiqknvgkfntss-----klts--lneeytknkteyeeaaqdaivkei--vnissgyv 589
Qy 292 RCIPMIFLNAPQGELENAFYFE-----EKGFGKI 321
Db 590 e--pmqtindvlag-ldavsvfahvngapypvyrpallekqggri 632

RESULT 15

AAR75411

ID AAR75411 standard; Protein; 934 AA.

XX AAR75411;

XX 14-NOV-1995 (first entry)

XX

```
DE Human MSH2.
XX
KW MSH2; mutator gene; DNA mismatch repair;
KW hereditary non-polyposis colorectal cancer.
XX
OS Homo sapiens.
XX
PN WO9515381-A.
XX
PD 08-JUN-1995.
XX
PF 02-DEC-1994; 94WO-US13805.
XX
PR 02-DEC-1993; 93US-0160295.
XX
PA (UYJO ) UNIV JOHNS HOPKINS.
XX
PI De La Chapelle A, Kinzler KW, Vogelstein B;
XX
DR WPI; 1995-215261/28.
DR N-PSDB; AAQ87269.
XX
PT New CDNA from human MSH2 gene encoding DNA mismatch repair enzyme -
PT its mutants causing hereditary non-polyposis colorectal cancer and
PT derived proteins useful in cancer prevention or treatment and for
PT diagnosis and screening.
XX
PS Disclosure; Page 37-38; 60pp; English.
XX
CC CDNA from human colon cancer cells was amplified using primers
CC previously used to isolate the yeast MSH2 gene from homology with
CC Muts. The insert in isolated clone PNP-23 was used to screen
CC CDNA libraries, and positive clones were used in a chromosome
CC walking procedure to identify the entire coding sequence (given
CC in AAQ87269) of the human MSH2 gene.
XX
SQ Sequence 934 AA;

Query Match 5.0%; Score 102.5; DB 16; Length 934;
Best Local Similarity 22.3%; Pred. No. 0.81;
Matches 77; Conservative 45; Mismatches 131; Indels 93; Gaps 17;

QY 18 GHMQVTQSVNQNDNMNL--DHLGVIE-----HDLFMEAHPILTICKKWI 62
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 63 NSFKEFRNMYKGYYSRPDKLDCFYKYYCLNKLINLL-IKEKPD-----LILLTPPTV 116
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 64 -----fqrqaanlqdcyrlggnglnpvnigalekhegkhqkillavftpl 440
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYVATKTKQDFIDVGIDPSTVKV 174
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 175 TGI--PIDNKEE--TPINQKQWLNLDNLDPKQITLMSAGAFVSGKGFDTMTDILAKSAN 231
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 482 reImndlekkmqstlisaar---dlglpdkgkldssaqqgyvfrvtckeekvlrnkn 538
QY 232 AQVVMICGSKSELKSLTAKFKLTRMYLILGYTKHNEWMASQMLITKPGGITTEGFA 291
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 539 fstvdikngvkkftns-----kts--lneeytknkteyeeaqdalvkel--vnissgyv 589
QY 292 RCIPMIFLNAPQGELENAFYFE-----EKGFCKI 321
Db | | | | | : | : | | | | | | | | | | | | | | | |
QY 590 e--pmqtlndvlaq-ldavvsfahvsnagpyvyrpailekqgqri 632
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 09:00:19 ; Search time 28.15 seconds

(without alignments)

1058.055 Million cell updates/sec

Title: US-09-668-788-4

Perfect score: 2048

Sequence: 1 MVTQNKILLITGSGFGNHM.....SSQPOBIYGVPLIARFFVK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	591.5	28.9	382	C69935	cell wall synthesis
2	309	15.1	464	C84499	probable monogalac
3	299	14.6	411	F75439	probable cell wall
4	296.5	14.5	468	T52269	1,2-diacylglycerol
5	284	13.9	533	T05092	probable 1,2-diacy
6	283	13.8	525	T10478	probable 1,2-diacy
7	245	12.0	373	B69860	conserved hypothet
8	230	11.2	374	E83894	hypothetical prote
9	136	6.6	363	F70195	UDP-N-acetylglucos
10	127.5	6.2	1157	S38160	NUP133 protein - y
11	124	6.1	363	E83970	UDP-N-acetylglucos
12	123.5	6.0	344	C70401	phospho-N-acetylmu
13	122	6.0	447	S35481	SRPM54 protein - M
14	121	5.9	359	T44337	hypothetical prote
15	117.5	5.7	353	C64561	transferase, pepti
16	116.5	5.7	353	G71852	udp-n-acetylglucos
17	116.5	5.7	357	E86823	peptidoglycan synt
18	115.5	5.6	398	E64456	hypothetical prote
19	113.5	5.5	765	S70962	regulatory protein
20	113	5.5	950	A71655	hypothetical prote
21	112.5	5.5	383	E70156	lipopolysaccharide
22	111.5	5.4	363	JC1275	phospho-N-acetylmu
23	109.5	5.3	531	A35343	glucuronosyltransf
24	109	5.3	354	D84955	hypothetical prote
25	108.5	5.3	363	C72590	probable hexosyltr
26	108.5	5.3	495	T28722	hypothetical prote
27	106	5.2	274	S60880	phosphorylation-ac
28	106	5.2	339	B72402	UDP-N-acetylglucos
29	106	5.2	390	A86065	hypothetical prote

30 106 5.2 530 2 A48633 glucuronosyltransf
31 106 5.2 533 2 T27589 hypothetical prote
32 106 5.2 1082 2 T43990 hypothetical prote
33 104.5 5.1 529 2 A24600 glucuronosyltransf
34 104 5.1 1196 2 JQ1467 toxin, nontoxic co
35 104 5.1 1196 2 S46430 botulinum neurotox
36 103.5 5.1 484 2 E64432 spore coat polysac
37 102.5 5.0 534 2 T19944 hypothetical prote
38 102.5 5.0 934 2 I64819 DNA mismatch repai
39 102 5.0 389 2 E65182 bacteriophage N4 a
40 102 5.0 398 2 F40511 hypothetical prote
41 101.5 5.0 368 2 F70346 mannosyltransferas
42 101.5 5.0 474 2 A64691 type III restricti
43 101.5 5.0 535 2 T57961 glucuronosyltransf
44 101.5 5.0 1248 2 A47445 reverse gyrase - S
45 101 4.9 530 2 T19365 hypothetical prote

ALIGNMENTS

RESULT 1

C69935

cell wall synthesis homolog yfp - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: C69935

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrazi,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtil

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A:Reference number: A69580; MUID:98044033

A:Accession: C69935

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-382 <KUN>

A:Cross-references: GB:299115; GB:AL009126; NID:g2634478; PIDN:CAB14110.1; PID:ell836

A:Experimental source: strain 168

C:Genetics:

A:Gene: yfp

Query Match

Best Local Similarity 28.9%; Score 591.5; DB 2; Length 382;

Matches 136; Conservative 77; Mismatches 131; Indels 37; Gaps 10;

Qy 5 NKKILLITGSGFGNHMQTSIVQNLNDMLDHLSEHDLFMEAHPILTSTCKKWIINS 64

Db 4 NKRVLILTANYGNHGVQAKTLYEQCVRLGFQHTV--SNLYQSPNPVSEVTQYLYLKS 61

Qy 65 FKYPNMYKGYYSRDPKDKCFYK-----YYGL-NKLIINLLIKE-KPDLILLTFPTPM 117

Db 62 FSIQKQFVRLFYG-----VDKIYKRNKRNFIYFKMGKRLGELVDEHQDIIINTFPPIVV 117

Qy 118 SVLTQEQNINPVATVMTDYRLHKNWITPSTRYVYVATKQDFIDVIGDIPSVKVTGI 177

Db 118 PEYRRRTGRVITPTNMTDFCLHKIWHENVDKYVATDYVYKELLETGTHPSNVKITGI 177

Qy 178 PIDNKFET-----PINOKQWLDNNDPDKOTILMSAGAFGVSGFGDFMTDILAKSANA 232

Db 178 PIRQFESMPVGYIKKY-----NLSNPKVLLIMAGAHGVLNKVELCEN-LVKDDQV 231

Db	370	ADALYDKNLGEAVGVVLICGRNKKLQSKLSSLDWKIP--VQVKGPITKMEECMGACDCI 427
QY	278	ITRPGGITTTEGFARCIPIFLNPAPQOELENAFYEEKGFGRKIADTPPEAKIVASLTN 337
Db	428	ITRAGPTIAEAMIRGLPITLNGYIAGAGNYPYVYVENCGRFSGFKSISKIVADWFG 487
QY	338	GNEOLTNNKIMTEODKTKYATQ---TICRDLDLI-GHSSQRP 376
Db	488	---PASKELETMSONALRAKPEAVFKIVHDMHIELVRKKNSLP 528

5

TI0478
 probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) pre-
 N/Alternate names: monogalactosyldiacylglycerol synthase
 C/Species: Cucumis sativus (cucumber)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: Tl0478
 R/Shimajima, M.; Ohta, H.; Iwamatsu, A.; Masuda, T.; Shioi, Y.; Takamiya, Y.;
 Proc. Natl. Acad. Sci. U.S.A. 94, 333-337, 1997
 A/Title: Cloning of the gene for monogalactosyldiacylglycerol synthase and
 A/Reference number: Z17042; MUID:97144442
 A/Accession: Tl0478
 A/Status: preliminary;
 A/Molecule type: mRNA
 A/Residues: 1525 <SHI>
 A/Cross-references: EMBL:U62622; NID:g1805253; PIDN:AAC49624.1; PID:g1805253
 A/Experimental source: cv. Aonagaibai; 5 day old seedlings

5

A;Genome: nuclear
C;Function:
A;Description: catalyzes the formation of monogalactosyldiacylglycerol, a
A;Pathway: membrane lipid biosynthesis
C;Keywords: chlroplast; glycosyltransferase; hexosyltransferase
F;1-103/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;104-525/Product: 1,2-diacylglycerol 3-beta-diacylglycerol #status
F;104-525/Product: 1,2-diacylglycerol 3-beta-diacylglycerol #status

Query Match 13.8%; Score 283; DB 2; Length 525;
Best Local Similarity 25.3%; Pred. No. 1.le-13;
Matches 105; Conservative 65; Mismatches 185; Indels 60; Gaps

QY 6[|] KKLIITSGFGNGHMVQTQSIVNQLNDMNLDHLSVEHDLFMEAHPILTICKKWYINSF 65
 :|::||: |::| |::| |::| |::| |:
Db 136 KRVILMSDTGGGHRASAEAKAAFEENFYGVFTDLWTHTP-----WPFNQI 187

QY 66 KYRN-----WKGPYYSRDPKL----DKCFKYKYLKNLNLIKPKDPLILLTFP-- 113
 |||::|||::|||::|||::|||::|||::|||::|||:
Db 188 PRSNEFLVKCTLKWKYYYVTAPKVTHOSFAETSEIAEVAAGLWKYPDDIIISVPDI 247

Qy 114 ---TPVMSVLTEQFNINIPATVMTDY-RLHKHWITPSTRYVYATKETKQDFIDVGIDP 169

Db 248 MQHVPILIRSKGLLNKIVITVTVDLSTCHPTWFHKLVTRCYCPSTEVAKRALTAGLQP 307

[illegible]

7

NE3011 /
B69860

conserved hypothetical protein ykoN - Bacillus subtilis

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: B69860
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta
C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, B.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: B69860
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Cross-references: GB:Z99110; GB:Z99111; GB:AL009126; NID:G2633699; PIDN:CAB13208.1; P
A:Experimental source: strain 168
C:Genetics:
A:Gene: ykoN

Query Match 12.0%; Score 245; DB 2; Length 373;
Best Local Similarity 23.5%; Pred. No. 4.9e-11;
Matches 88; Conservative 75; Mismatches 168; Indels 44; Gaps 14;
QY 6 KKLIIITG-SFGNGHMQVTSIVNQLNDMLDHLVIEHDLFMEAHPLTISICKKWINS 64
DB 2 KNLIFFPLSITGHHVADALAELESQ----LAAEKIDIFSRYRLEKLSVVALKW 58
QY 65 PKYFNMYKGYSPDKL-----DKCFYKYGL--NKLINLLIKERPDILLT--F 112
DB 59 IQYFKTYGIY----RLLAGCFQHDKRYFMVECVFTQQRHLQEKQPDIAFCTHAL 113
QY 113 PTPVMSVTEQFNINIPVATWTDRLHKNMITYSTRYVATKQDFIDVGDISTV 172
DB 114 PSYLNRLKPEYP-MLTVNVTDFVFNQLWGRKNIDYHFPVSTEVKQLISEGIDQNNI 172
QY 173 KVTGPIDNKPTETPINOQWLNDLNDPKOTILMSAGAFGVSGFDMTDILAKSANA 232
DB 173 YLTGIPVHQNFEME-----SADTLQHHPPYTIITGSGMGV-GILKWVQV-LSPGKI 224
QY 233 QVVMICGSKELKRLS-TAKFKLTRMYLILGYTKHNMWMASSOLMITKPGGITTEGFA 291
DB 225 LYKILCGRNEKLYSVKSLHPLIEAIPYLHSAKEMRLYEQATGIMTKPGGVITSECLQ 284
QY 292 RCIPMIFLNPAPQGLENAFTFEKGFQKIAD---TPEEAIIKIVASTNNEQL----- 342
DB 285 KRLPVFIYHALPGQEMNLNLLHERKL--VTDNRNWDMDQKAEEXIAAFQSQEQMKYK 342
QY 343 --TNMISTMEQDKIK 355
DB 343 HVNGYLGEMSDRKIK 357

RESULT 8
E83894
hypothetical protein BH1957 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83894
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314

A:Accession: E83894
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <STO>
A:Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05676.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1957

Query Match 11.2%; Score 230; DB 2; Length 374;
Best Local Similarity 21.8%; Pred. No. 6.7e-10;
Matches 83; Conservative 82; Mismatches 165; Indels 50; Gaps 14;
QY 9 LIITGFGNGHMQVTSIVNQLNDMLDHLVIEHDLFMEAHPLTISICKKWINSFKYF 68
DB 6 LIFSASIGNGHNAQAKAL--OVFQKNGYQPEI-IDTFYLSLPAHKFMTSVNLLKVG 62
QY 69 RNMKGFFYSRPDK-----LDRCFYKYGLNKLINLLIKERPDILLTFF--TPVMSVL 120
DB 63 PRIWQKIYF-QAEKYPLEFLDQ--FATPFVESLHATVKSNCSESLVSTHPPFVTAFLVRL 119
QY 121 TEQFNINIPVATWTDRLHKNMITYSTRYVATKQDFIDV-GIDPSTVKVGTGPI 179
DB 120 KSKKOLLPLTYITDFVLHPAYLRPEIDGYFTSDPNF-TDFAKLNNSDDRFPTGTGPI 178
QY 180 DNKFTETPINQKWLIDNN--LDPDKOTILMSAGAFGVSGFDMTDILAKSANAQVMI 237
DB 179 PNL-ESIDQPKVNRNDLGLDQPKVLLIAGGGIGLT-NYAQVIRALECLPEPIQLLCL 235
QY 238 CGSKELKRLSATAKFKLTRMYLILGYTKHNMWMASSOLMITKPGGITTEGARCIPI 297
DB 236 IGNYQVKEKI-SRIKSKHELKVIETDFKFLYLKASDAILSKAGGLTMAESLVCEPTII 294
QY 298 FLNAPAQGLENAFY-----FEKGFQKIADTPPEAKIVA 333
DB 295 IHQVPQHEHNAKFLIDAGALRVKSKKEIPTIKRVLYEEACFGPMIENARKLKPN 354
QY 334 SLTNGNEQLTNMISTMEQDK 353
DB 355 A----NEIVEQMLLVKEQ 370

RESULT 9
F70195
UDP-N-acetylglucosamine--N'-acetyluranyl-(pentapeptide) pyrophosphoryl-undecaprenol N
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 24-Nov-1999
C:Accession: F70195
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: F70195
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-363 <KLE>
A:Cross-references: GB:AE001176; GB:AE000783; NID:g2688699; PIDN:AAC67113.1; PID:g268
A:Experimental source: strain B31
C:Superfamily: murG protein

Query Match 6.6%; Score 136; DB 2; Length 363;
Best Local Similarity 21.2%; Pred. No. 0.0082;
Matches 86; Conservative 68; Mismatches 155; Indels 96; Gaps 18;
QY 5 NKKILITGSPFGNGHMQVTSIVNQLNDMN-----LDHLVIEHDLFMEAHPI--LTS 55
DB 3 NKKIIFGTGGTGHHVFGPSIGIIQKLEFDEINEIFFWIGKKNSTEELIKRQDNKIFSI 62

285 ITTNK-----GIFQTWOLGATNSHPHTKLDIVNI-YEAILESQDLYPFAHGTLKWDSH 337

Qy 51 PI-----LTSI-----CRK-WYINSFKYFRNMVKGYFSRDKLDKCFKYKYGLNKLI 97

Db 338 PLQDESSQLFSLSIYDSSCNETYYILSTIIFDSSNSF-----TIFSTYRLNTFM 387

Qy 98 NLL--IKEKPDL-----ILLTPTPVMSVLTEQFNINIPVATMTDYRL 139

Db 388 ESITDTRFKPIFQPMENANDTNEVTSILVMFPNAV--VITQ---VN---SKLDSYSM 439

Qy 140 HKNWITPVSTR-----YVVAIKE--TKQDFI-----DVGIDSTV- 172

Db 440 RKKEDIVSLRNDIIDIGSGYDSKSLVLRKQMGVLQFFVKENEETNSRPEVGFVASHVD 499

Qy 173 -----KVTGIPIDNKFPETPINOKQWLIDNNLPDKOTILMSAGF-----GV 214

Db 500 QAVYFSKINAPIDFNLPETSLDQESIEHDLKLTSEIPIHSGNKVYIPPMNLTLGQHLV 559

Qy 215 SKGFDTWITDLAKSANAQVVMICGSKSELKRSITAKFKLTRMYL-----259

Db 560 RKEEFQNLFTVAKNFYKI-----SPELKLDLIERKEFELNCCIKFNSTIROSVDLNDI 613

Qy 260 ----ILGYTKHNMWMASSOLMITKPGGITTEGFARCIPMIFLNPAPGOELENAFYEE 315

Db 614 WEKTLNSVNLTONHLLTKTVINSPD-----VFPVIF-----KQFLNHVVF-- 655

Qy 316 KGFGKIADTPEEAIKIVASTNGNEQL--TNMISTWEDQKI-KYATQTCRDLDDLIGH 372

Db 656 -----VLFPSQNONFKLVNTNLNLCFYDGILEGEKTIIRYELLE-----696

Qy 373 SQPOEI-YGKVPLYARF 388

Db 697 -DPMEVDTSKLPWFINF 712

RESULT 11

E83970

UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophos murG [Impo

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

C:Accession: E83970

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.;

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus h

A:Reference number: A83650; MUID:20263314

A:Accession: E83970

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA806284

A:Experimental source: strain C-125

C:Genetics:

A:Gene: murG

C:Superfamily: murG protein

Query Match 6.1%; Score 124; DB 2; Length 363;

Best Local Similarity 19.8%; Pred. No. 0.066;

Matches 77; Conservative 82; Mismatches 149; Indels 80; Gaps

Qy 9 LIITSGNGNHMQVTSIGNOL--NDNMLDLHSV-----IEHDL-----FMEAHPILT 54

Db 3 IVVSGGTGTHIYPALAFINEMKKRDLVDLVICTERGLESIVREGIPTQTH--IT 60

Qy 55 -SICKWYINSFKYFRNMVKGYFSRDKLDKCFKYKYGLNKLINLIKE-KPDLILLTTP 113

Db 61 -GFORKLSMENVKTVVRLRG-----TKRAKALLNEFKPDVVGITGG 101

Qy 114 TPVMSVLTEQFNINIPVATMTDYR---LHKNWITPVSTRYVVAKEFKQDFIDVGIDP 169

Db 102 YVCGPVVYAAAKLUP--TVIHEQNSVPGVTNFKFLSRVDRIAICFKAEEAF-----PK 154

Qy 170 STVKVTGIPIDNKFETPINOKQWLIDNNLPDKQTTILM---SAGAFGVSKGFDTMTDIL 226

Db 155 NKVVFTGNPRASEVMSG-NREGLRSLGKIPKPKTKVLIIVGGSRGARPINDAFMSILSDVK 213
QY 227 AKSANAQVVMICG-----KSKELKSLTAKFKLTRYMLILGYTKHNEWMASQMLMTKP 281
Db 214 AKP--YQFYVVTGVHYERVQSMKSIGOPENV---IVQPIHNPDPVLSAVIDIVARA 267
QY 282 GGTITTEGFARCPIMFLNPAP---GOLENAFYFEKGFQGIADTPEAKIVASLTN 337
Db 268 GATTLEITAGLPSILI-PSYVYNNHGEKNAALSCK-----DAAILRESELT 317
QY 338 GNEQLTN-----MISTMEQDKIKYATQTI 361
Db 318 GDRLEDIDDDINVTGRLDAMKQAAKAL 345
RESULT 12
C70401
phospho-N-acetyluramoyl-pentapeptide transferase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: C70401
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: C70401
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-344 <AQF>
A:Cross-references: GB:AE000727; NID:92983623; PIDN:AAC07193.1; PID:92983625; GB:AE00069
A:Experimental source: strain VF5
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 6.0%; Score 123.5; DB 2; Length 344;
Best Local Similarity 20.7%; Pred. No. 0.067;
Matches 80; Conservative 72; Mismatches 131; Indels 103; Gaps 19;
QY 2 VTQNKILLITGSGNGHMQVTSIVNQLNDMLDHLVIEHDLFMEAHPIILTSICKKWY 61
Db 23 VLKKEVKVFGSKRG-----IEYELKDL-----INTEKFLDVEPLRE----- 60
QY 62 INSFKPRNMVYKGFYSRPDKLDCFKFYGLNKLINLLIKERPDILL-----TFTPTVM 117
Db 61 -----RNFIQ-----KL-KAIWKFLKAOEEINEFLKEDYRALIFGGYASLPLGIN 104
QY 118 SVLTEQPNIPVATVMTDYRLHKNWITPYSTRYVATKE-----TKQDFIDVGID 168
Db 105 TVLRK-----ELFIHEQNSIPSKTNKILSKAKKVLITFNKTRFF----- 146
QY 169 PSTVKVTGIPIDNKFPPINOKWLDNLDPKQTLMSAGAFVSGKGFDTMITDLAK 228
Db 147 PEGVRV-GLPIRKLKLPKPKVKKRFGLEPKITVLI-----FGSQG--ALFLNELAR 199
QY 229 SANA-----QVVMICG--SKELKSLTAKFKLTRYMLILGYTKHNEWMASQMLMTK 280
Db 200 DLKSVLPKFEQVILLTKIHYEFKNLEGEKFR-----VMPSLDGLIYSASDVISR 253
QY 281 PGGTITTEGFARCPIMFLNPAPQOLENAFY-----FEKGFQGIADTPE-----EAI 329
Db 254 AGACTINELSHFGVPSVFV-PYPVAVDDHQFYNAKETEKLGGLVLRQEEAKPKDVLKAL 312
QY 330 K-IVASLTNGNEQTLTNMISTWEQDKI 354
Db 313 KEIVKNLERYSENKPKFAEGAEERM 338
RESULT 13
S35481

SRPM54 protein - Mycoplasma mycoides
C:Species: Mycoplasma mycoides
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
C:Accession: S35481; S27591
R:Samuelsson, T.
Nucleic Acids Res. 20, 5763-5770, 1992
A:Title: A Mycoplasma protein homologous to mammalian SRP54 recognizes a highly conse
A:Reference number: S35480; MUID:93087189
A:Accession: S35481
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <SAM>
A:Cross-references: EMBL:M91593; NID:gl50208; PIDN:AAA25441.1; PID:gl50210
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: signal recognition particle 54K protein

Query Match 6.0%; Score 122; DB 2; Length 447;
Best Local Similarity 22.9%; Pred. No. 0.13;
Matches 80; Conservative 52; Mismatches 143; Indels 74; Gaps 16;
QY 13 GSGFGNGHMQ--VTOSIVNO-LNDMLND-----HLSVIEHDLFMEAHPIILTSICKKWYN 63
Db 4 GDFLSKRMQKSIKKNKSTLNEENIKETLKEIRLSLLEADVNEA-----AKEIIN 55
QY 64 SPKYFRNMVYKGFYSRPDKLDCFKFYGLNKLINLLIKERPDILLPTPTVMSVLTEQ 123
Db 56 NVK---QKALGYSISEGASAHQOMIKIVH-EELVNILKGNAPLDINKKPSVMMVGLQ- 110
QY 124 FNIINIPVATVMTDYRLHKNWITPYSTRYVATKEKQDFIDVGID---PSTVKVTGIPID 180
Db 111 -----GSKTTTANKLAYLLNKKKKVLLVGLDIYRPGAIEQL-VQLG 153
QY 181 NKFETPIINOKWLDNLDNLDPKQTLMSAGAFVSGKGFDTMITDLAKSANAQVVMICOK 240
Db 154 QKTNITQVFEK-----GKQDPVKTA--EQALEYAKENNFDWILDTAGRLQVDQVLM- 202
QY 287 TEGFARCPIMFLNPAPQOLENAFYFE-----KGFQGIADTPEPAIK 330
Db 261 STSYLTKLPKIFGEGEGYNALAAFPYKPMADLMGMDIETILFERAVE 309

RESULT 14
T44332
hypothetical protein wble [imported] - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44332
R:Yamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nair, G.B.; Takeda,
Gene 237, 321-332, 1999
A:Title: The genes responsible for O-antigen synthesis of Vibrio cholerae O139 are cl
A:Reference number: Z22749; MUID:99453293
A:Accession: T44332
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-359 <VAM>
A:Cross-references: EMBL:AB012957; NID:g4115688; PIDN:BAA33636.1; PID:g3721686
A:Experimental source: strain O22
C:Genetics:
A:Note: wble
C:Superfamily: probable hexosyltransferase ytxn

Query Match 5.9%; Score 121; DB 2; Length 359;
Best Local Similarity 19.3%; Pred. No. 0.11;
Matches 79; Conservative 76; Mismatches 139; Indels 116; Gaps 19;
QY 10 IITGSGNGHMQVTSIVNQLNDMLDHLVIEHDLFMEAHPIILTSICKKWYNSEKRYFR 69

Db 11 VITSDSGG---AENLVSNMLDSD-----GLFLQ-YAI-----YFS 42
QY 70 NMYKGFYSPDKLDRKCFYKYG-----LNKLINLLKEKPDILLTFPT-PVM 117
Db 43 SFISG--YRRPNTI-----YFGSNRSLNIFKLRKTIKMLIEKHGELIVHAHLTWPLF 94
QY 118 SVLTEQFNINIPVATVMTDRLHKNWITPYSTRYVATKTKODFID----- 164
Db 95 YVALGLENLVNLV-----YEHST-----FNKRRKYPFLKFFRELVIYSRSRI 138
QY 165 VGIDPSTVKVTGIPIDNKFETPINOKOWLIDN-----NLDPKQOTILMSAGAFG 213
Db 139 VGISNSVSKSIEVWLGSSFLSRI-----VTINNGARFYSPKCRHNLELSKKILLVSGSLK 194
QY 214 VSKGFDMITDI--LAKSANAQVVMICGSKELKRLSTAKFKLTRMYLLILGYTKHNEWM 271
Db 195 PLKGFDTIKSLKITQYIDKYITVGTGSPSYTLKEISLDNHVDHLVDFVGSNELEKFY 254
QY 272 ASSQLMIT-----KPGGITITEGFARCIPIMF-----LNAPAGOELENAF----YFEKGF 318
Db 255 HEADILLIPSHWEGFGLSAGEHSTGLPVIASNVVDGLNEVSVKTLASSILVENFLDENS 314
QY 319 GKIADPPEAKIVASLTNGNEQLTNMISTMEQDKIKYATQTCICRDL 368
Db 315 ---ADAILEMKNKLLSS-----DYFNMSIESKQOAYKFSDFKMFKEYTDL 355

RESULT 15

C64664
transferase, peptidoglycan synthesis - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Feb-2000
C:Accession: C64664
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: C64664
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-353 <TOM>
A:Cross-references: GB:AE000621; GB:AE000511; NID:g2314301; PIDN:AAD08196.1; PID:g231430
C:Superfamily: murG protein

Query Match 5.7%; Score 117.5; DB 2; Length 353;
Best Local Similarity 20.6%; Pred. No. 0.2;
Matches 81; Conservative 59; Mismatches 169; Indels 85; Gaps 20;

QY 11 ITGSFGNGHMVQTSIVNQLNDNLDHL---SVIEHDLFMEARHPILTSICKKWIINSFKY 67
Db 5 LTGGTGGHLSIAKALAELEKOGIEAIYLGSTYGQD-----KEWFENS-PL 50
QY 68 FRNMYKGFYSPDKLDRKCFYKYG---LNKLINLLIKE--KPDILLTF-----PT 114
Db 51 FSERY---FNTQGVNKSFFKIGSLFLQAKAFAKAKELKKHQITHITSVGFSGNPA 107
QY 115 PVMSTVEQFNINIPVATVMTDRLHKNWITPYSTRYVATKTKODFIDVGI-DPSTVK 173
Db 108 SFASLLNK-----IPL-----YIHEQNAIKGSLNRY--LSPRAKAVFSYAFKDKGNHV 154
QY 174 VTGIPIDNKF-----ETPINQKQWLIDNLDPKQOTILMSAGAFVS--KGFDTMITDI 225
Db 155 LTSYPVQNAFFDPAFRTTEKH-----ILFLGSGQAKAINEFALLNAPK 199
QY 226 LAKSANAQVVMICG-KSKELKRLSTAKFKLTRMYLLILGYTKHNEWMASQLMITKPGGI 284
Db 200 LTKQ-GIKITHICGPNSEQVRFYQELGLLDLKIELFAFHNNITEIMHRADLCVSRAGAS 258

QY 285 TITEGFARCIPIMFILNAP-----GOELENAFYEEKGFGKIAD-----TPEAKIIVASLT 336
Db 259 SVWELCANGLTTFI-PYPFASNNHQYVNVLEFEKENLCYVVPONELLPKKLFVIRKLN 317
QY 337 -----NQEQLTNMISTMEQDKIKYATQTCICRDL 366
Db 318 QKDDQGNKNTTISNQLQKQKIADGAKTIETIL 351

Search completed: June 29, 2001, 09:00:20
Job time: 182 sec

3

KW Membrane; Peptidoglycan synthesis.
SQ SEQUENCE 363 AA; 39916 MW; 5178BCC95264BE69 CRC64;

Query Match 6.68; Score 135; DB 1; Length 363;
Best Local Similarity 20.68; Pred. No. 0.0035;
Matches 83; Conservative 75; Mismatches 161; Indels 84; Gaps 16;

QY 9 LIITGSGFGHGMQVTSVNLQNDMLDHLVIEHDFMEAHPILTSTICKKWIYNSPKY- 67
DB 3 ILVGGGTGGHYA-----LSFVEH-----VKKEAPATEFLYV 36

QY 68 -----FRNM-YKGFYYSRDPKLDKCFYKY-GLNKLINLLIKEKPDLI 108
DB 37 GTENGLESQIVPKAKIPEKTIKIQFKRSLSPQNFKTYLTLTSTINKAKKIREFQPDV 96

QY 109 LLTFTPTVMSVLTQFNINIPVATMDYRLHKNWITPYSTRYV-----VATETK 159
DB 97 IGTGGYSGAVVYAAHOLKIP--TIIHQNSIPGMWNKFLSRVYDKIAICPPDVASFPPK 154

QY 160 QDFIDVGDIPSTVKTGIPDNKFPETPNQKWLIDNNLDKOTILM---SAGAFGVSK 216
DB 155 EKT-----LTGNPRQEVVT-VEKSAILSEFGLDPAKTVLVFGSGRALKINQ 203

QY 217 GFDWITDILAKSANAQVVMICGKS--KELKRSUTAKFKLTRYLLIILGYTKHNMENMASS 274
DB 204 AFEQAQP--LFEEREYQVLYASGERYQELQESLKLSEKLTNISVOPYIDKMYEVNANT 261

QY 275 QLMWTKPGIITITGFCARCPIMFLNAP-----GOELENAYFEEKG-FGKIADTPPEAI 329
DB 262 DLWGRAGATSIAEFTALGLPAILI-PSYVNDHQTKNAQSLVKVGAVERIPDAELTGA 320

QY 330 KIVASLTN---GNEQLTNMISTMEQDKIKYATQICRDLDLI 369
DB 321 RLVAADIDLLNNEKQOMATASKGERIPDASDRLYQVQVTKIV 363

RESULT 4
N133_YEAST ID N133_YEAST STANDARD; PRT; 1157 AA.
AC P36161;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NUCLEOPORIN NUP133 (NUCLEAR PORE PROTEIN NUP133).
GN NUP133 OR YKR082W OR YKR402.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUXJR;
RX MEDLINE=95112817; PubMed=7813444;
RA Doye V., Wepf R., Hurt E.C.;
RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
RNA transport and nuclear pore distribution.";
RL EMBO J. 13:6062-6075(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262327; PubMed=8203164;
RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
Remacha M.A., Revuelta J.B., Ballesta J.P.G., Jimenez A., del Rey F.;
RT "The complete sequence of an 18,002 bp segment of Saccharomyces
cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
and six new open reading frames.";
RL Yeast 10:231-245(1994).
CC -1- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE
DISTRIBUTION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80066; CAA56372.1; -
DR EMBL; 227116; CAA81633.1; -
DR EMBL; 228307; CAA82161.1; -
DR PIR; S38160; S38160.
DR PIR; S39123; S39123.
DR SGD; S0001790; NUP133.
KW Nuclear protein; Transports; Transmembrane.
FT TRANSMEM 217 233 POTENTIAL.
FT SIMILAR 98 420 TO NUP120 (AA 434-763).
SQ SEQUENCE 1157 AA; 133319 MW; C8BDBB7D709C5C08 CRC64;

Query Match 6.28; Score 127.5; DB 1; Length 1157;
Best Local Similarity 20.98; Pred. No. 0.063;
Matches 104; Conservative 67; Mismatches 147; Indels 179; Gaps 27;

QY 2 VTQNKKILITGSGFGHGMQVTSVNLQNDMLDHLVIE--HDLF-----MEAH 50
DB 285 ITTNK-----GIFQTWQLSATNSHPKLDVNI-YEAILSLQDLYPFAHGTLKINDSH 337

QY 51 PI-----LTST---CRK-WYINGFKYFRNMKGFYYSRDPKDKCFYKYGKLNKLI 97
DB 338 PLQDESSQLFLSSYDSSCNETYVILSTIIFDSSNSF-----TIFSTRVLTFTM 387

QY 98 NLL--IKEKPDL-----ILLTFTPTVMSVLTQFNINIPVATMDYRL 139
DB 388 ESITDTRKFKPKIFPQMANANDTNEVTSILVMPNAV--VITQ---VN---SKLSSYSYM 439

QY 140 HKWITPTSTR-----YVATKE--TKODFI-----DVGIDPSTV- 172
DB 440 RRMWEDIVSLRNDIDITGSGYDSKSLYVLTQKMGVLQFFVKENETNSKPEVGFVKSHVD 499

QY 173 -----KVTGIPDNKFPETPNQKWLIDNNLDKOTILMSAGAF-----GV 214
DB 500 QAVYFSKINANPIDFNLPPEISLDQESIEHDLKLTSEIFHSNGKYIPPMMLTLGQHLVS 559

QY 215 SKGFDWITDILAKSANAQVVMICGSKELKRSUTAKFKLTRYLLI----- 259
DB 560 RKEFFQNFQVAKNENYKI-----SPELKLDLIEKFEILNCCIKFENSIIRQSDVLNDI 613

QY 260 ----ILGYTKHNMENMASSQLMTKPGIITGFCARCPIMFLNAPGOELENAYFEE 315
DB 614 WEKTLNNTQNEHLTTKTIVINSPD-----VFPVIF-----KQFLNHHVVF-- 655

QY 316 KGFQKIADTPPEAKIVASLTNGNEQL--TNMISTMEQDKI-KYATQICRDLDLIGH 372
DB 656 -----VLFPSQNFKNLNTNLNLCFYDGLIEGEXTIRYELLEL----- 696

QY 373 SQPOEI-YGKVPLYARF 388
DB 697 -DPMEVDTSKLPWFINF 712

RESULT 5
MURC_STRPN ID MURC_STRPN STANDARD; PRT; 352 AA.
AC Q9ZHA9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLURAMYL- (PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC
DE TRANSFERASE).
GN MURC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

SR54_MYCMY STANDARD; PRT; 447 AA.

ID AC Q01442; 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG).

GN FFR OR SRP54.

OS Mycoplasma mycoides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Entomoplasmatidae.

OX NCBI_TaxID=2102;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93087189; PubMed=1280809;

RA Samuelsson T.B.;

RT "A Mycoplasma protein homologous to mammalian SRP54 recognizes a highly conserved domain of SRP RNA.";

RL Nucleic Acids Res. 20:5763-5770(1992).

CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE RIBOSOMES (BY SIMILARITY).

CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA MOLECULE AND PROTEIN FFH (BY SIMILARITY).

CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL SEQUENCE.

CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M91593; AAA25441.1; -

DR PIR; S27591; S27591.

DR PIR; S35481; S35481.

DR HSP; O07347; LFPH.

DR InterPro: IPR000897; -

DR Pfam: PF00448; SRP54; 1.

DR PROSITE: PS00300; SRP54; 1.

KW Signal recognition particle; GTP-binding; RNA-binding.

FT DOMAIN 1 295 G-DOMAIN.

FT NP_BIND 296 447 M-DOMAIN.

FT NP_BIND 108 115 GTP (BY SIMILARITY).

FT NP_BIND 190 194 GTP (BY SIMILARITY).

FT NP_BIND 248 251 GTP (BY SIMILARITY).

SQ SEQUENCE 447 AA; 50100 MW; 24A71128D4041D43 CRC64;

Query Match 6.0%; Score 122; DB 1: Length 447;

Best Local Similarity 22.9%; Pred. No. 0.045;

Matches 80; Conservative 52; Mismatches 143; Indels 74; Gaps 16;

QY 13 GSGFNGHMQ-VTQSIYVQ-LNDMNL-D-----HLSVIEHDLFMEARHPILTISICKKWIYN 63

DB 4 GDFLSKRMQKSIKKNMNSTLNEENIKETLKEIRLSLEADVNI EA-----AKEIIN 55

QY 64 SFKYFRNMYGFFYSRDKLDKCFYKYVGLNKLINLLIKEDPLDILITFTPTWMSVLTEQ 123

DB 56 NVK---QKALGGYISGASAHQMIKTVH-EELVNLGLKENAPLDINKKPSVVMVGLQ- 110

QY 124 FNINIPVATVMDYRLHKNHITPYSTRYVYATKTKQDFIDVGID---PSTVKVTGIPID 180

DB 111 -----GSGKTTTANKLAYLLNKNKKVLLVGLDIYRPAIQSL-VQLG 153

QY 181 NKETPTNOKWLDNNLPDKPTILMSAGAFGSKGFDPTMTDILAKSANAQVVMICGK 240

DB 154 QKTNTOVFEX-----GRQDPVKA--EQALEYAKENNFVVDLTAGRLQVDQVLM---- 202

QY 241 SKELRSITAKFKLTRMYL-----ILGYTKHNMENWASSOLMTK-----PGCITI 286

DB 203 -KELD-NLKKTSFNEILLVVDGMSGQEIINVTFNFDKLSGVVVTKLDGDARGGATL 260

QY 287 TEGFARCIPIFIPLNPAPQOLENAFYEE-----KGFGKIADTPPEAIK 330

DB 261 SISVLTPLPIKFIGEGSYNALAAFYKPRMADRLMGMDIETLFAVE 309

RESULT 8

UD17_RAT

ID AC Q64633; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE UDP-GLUCURONOSYLTRANSFERASE 1-7 PRECURSOR, MICROSOMAL (EC 2.4.1.17)

DE (UDPGT) (UGT1*7) (UGT1-07) (UGT1A7) (A2).

GN UGT1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE OF 1-286 FROM N.A.

RC STRAIN=WISTAR;

RX MEDLINE=95332265; PubMed=7608130;

RA Emi T., Ikushiro S.I., Iyanagi T.;

RT "Drug-responsive and tissue-specific alternative expression of multiple first exons in rat UDP-glucuronosyltransferase family 1 (UGT1) gene complex.";

RL J. Biochem. 117:392-399(1995).

RN [2]

RP SEQUENCE OF 287-531 FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=90274676; PubMed=2112380;

RA Sato H., Koiwai O., Tanabe K., Kashiwamata S.;

RT "Isolation and sequencing of rat liver bilirubin UDP-glucuronosyltransferase cDNA: possible alternate splicing of a common primary transcript.";

RL Biochem. Biophys. Res. Commun. 169:260-264(1990).

CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND ENDOGENOUS COMPOUNDS.

CC -1- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR -> UDP + ACCEPTOR

CC BETA-D-GLUCURONOSIDE.

CC -1- SUBCELLULAR LOCATION: MICROSOMAL.

CC -1- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES FOR MANY DIFFERENT UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMINAL DOMAIN AND A COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.

CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D38062; BAA07258.1; -

DR EMBL; M34007; AAA42312.1; ALT_TERM.

DR InterPro: IPR002213; -

DR Pfam: PF00201; UDPGT; 1.

DR PROSITE: PS00375; UDPGT; 1.

KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;

KW Multigene family; Microsome; Alternative splicing.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 531 UDP-GLUCURONOSYLTRANSFERASE 1-7.

FT TRANSMEM 48 503 POTENTIAL.

FT TRANSMEM 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
SQ SEQUENCE 531 AA; 59627 MW; BC791DCE724CA621 CRC64;

Query Match 5.8%; Score 119; DB 1; Length 531;
Best Local Similarity 18.3%; Pred. No. 0.096;
Matches 79; Conservative 70; Mismatches 123; Indels 160; Gaps 20;

QY 23 TOSIVNQLNDMLDHLVIEHDLFEAH-----PILTSICKWYINSFYRNMV 72
DB 76 TVSVHTQEDLNR-----EFKFFIDSQWTKQBSGVLPILTSQAQFFELFHSCHSLF 129
QY 73 KG---FYSRPKDKLCFY-----KYGLNKLI----- 97
DB 130 KDKLVEYLKQSSFDAVFLDPDVCGLTVAKYFSLPSVFSRGIFCHYLERGSCQSPSPS 189
QY 98 -----NLL-----IKEKPDLLILLTPTTPVMSVLTQFN 125
DB 190 VYVRPILKLTDMTKERWNLLSYMGHAFPCPFKATDIASEVLQTPV--TWTDLFS 247
QY 126 INIPVATVMTDYRLHKNWITPYSTRYVATKQDFIDVGIDPSTVKVTGI-----PI 179
DB 248 -PVSVMLLRTDTLE-----LPRVMPNVIIHGGINCHQRKPV 284
QY 180 DNKFETPINQKWLIDNNLDPKQILMSAGAFV---SKGFTMTIDILAKSANAQVVM 236
DB 285 SKFEFAYVN-----ASGEHGIWVFSLG--SMVSEIPEKKA-METAE 322
QY 237 ICGSKELKRSITAKFKLTRYMLILGYTKHNMENWASSQLM-----ITKPGGITITE 288
DB 323 ALGR---IPOTLLWRYTGRFNSLAKNT-ILVKWLPQNDLGHKARAFITHSGSHGIVE 378
QY 289 GFARCIPIMLFNPAPQOELENAFYFEERKFG-----KIADTPBEAIIKIVASLTNGNEQ 341
DB 379 GICNGVPMVM--PLFGQDMQNAKMETRGAGVTLNVLEMTADDLENALKTVINNKSYKEN 437
QY 342 LNNMISTMEQDK 353
DB 438 IMR-LSSLHKDR 448

RESULT 9
MURG_HELPJ
ID MURG_HELPJ STANDARD; PRT; 353 AA.
AC O25770;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGUCOSAMINE TRANSFERASE
DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE TRANSFERASE).
DE MURG OR HP1155.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
```

RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.B., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic-sequence comparison of two unrelated isolates of the human
 gastric pathogen *Helicobacter pylori*.";
 RL Nature 357:176-180(1999)
 CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
 CC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
 CC INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
 CC (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
 CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ235272; CAAL4963.1; -;
 DR InterPro; IPR001646; -;
 KW Transferase; Glycosyltransferase; Cell division; Cell wall;
 KW Membrane; Peptidoglycan synthesis.
 SQ SEQUENCE 353 AA; 39506 MW; 1773AD7E5C214EF9 CRC64;

Query Match 5.7%; Score 116.5; DB 1; Length 353;
 Best Local Similarity 19.8%; Pred. No. 0.086;
 Matches 79; Conservative 63; Mismatches 162; Indels 95; Gaps 20;
 QY 11 ITGSGFGHGMQVTSIVNQLNDMLHL---SVIEHLDFEAPHLISICKKWINGEFKY 67
 DB 5 LTGGGTGGHLSIAKALAELEKQIEAIIYLGSTYQD-----KENFENS-PL 50
 QY 68 FRNMYKGYFRPDKLCKFYKGLNKLINLLIKERPDL-----ILLTF----- 112
 DB 51 FSERY---FFNTQGVNKSFFK-----KIRSLFOAKAFAKAKELKQHITISVGGF 102
 QY 113 ---PTPVMVLTGFNINPVATVMTDYRLHKWITPYSTRYVATKTKQDFIDVGI-D 168
 DB 103 SAGPASFASLANK-----IPL-----YIHEQNAIKGLNRY--LSPKRAVFSYAFKD 149
 QY 169 PSTVKVTGIPIDNKF-----ETPINQKWLIDNLDPDKOTILMSAGATGVS--KGFD 220
 DB 150 KGNHVLTSYPQNAFFPHARTREIKH-----ILFLGSGQAKAINEFAL 194
 QY 221 MITDILAKSANAQVVMICG-KSKELKRSLTAKFKLTRMYLILGYTKHNMENWMASSQLMIT 279
 DB 195 LNAPKLTKQ-GIKTHICGPNSEQVRFYQELGLDKIELFAPHNNITEVMHRADLCVS 253
 QY 280 KPGGITITEGFARCPIMFLNPAP---GOELENAFYFEKGFCKIAD---TPEEAIKI 331
 DB 254 RAGASSWELCANGLPITIFI-PYPFASNHHQYINVLEFENLCYVVPQNELLPKLFVEY 312
 QY 332 VASLT---NGNEQLTNMISTMEQDKIKYATOTICRDL 366
 DB 313 IRKLNQDDQGNKLTWISAKLQKIAKDGAKTIETIL 351

RESULT 11
 Y511_RICPR STANDARD; PRT; 950 AA.
 AC Q92D36;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL PROTEIN RP511.
 GN RP511.

OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ235272; CAAL4963.1; -;
 DR InterPro; IPR001646; -;
 KW Pfam; PF00805; Pentapeptide; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 950 AA; 108612 MW; C047F8BCF063F715 CRC64;
 Query Match 5.5%; Score 113; DB 1; Length 950;
 Best Local Similarity 20.4%; Pred. No. 0.61;
 Matches 92; Conservative 69; Mismatches 156; Indels 134; Gaps 21;
 QY 4 QNKILLITSGFGHGMQVTSIVNQLNDMLHLVIE--HDLFMAHPILTSICKWKY 61
 DB 209 EKKLLKIF---NTHPDVKQLVNNINPN-----ILKRFNKLIFYKQELMTSLFKEVK 259
 QY 62 INSPKYFRNMVKGK-----YYSRPDK-----LDKCFYKY 91
 DB 260 AQSPFLOEHFESKIDKILDIITLLNKPDIKEIFDILNAPKGMISLEKALEWA 319
 QY 92 GLNKLINLLIKEK---PDIL-LTFPTVMSVLTEQFNIN---IPVATVMTDYRLHKWN 143
 DB 320 GDOQLKSFANNKTLIPNLMAIGIETNPVSQISITNEYNFQOQMLVIVGEVMSKPEIAHEI 379
 QY 144 ITPVSTRYVATKTKQDFIDVGI-DPSTVKVTGIPIDNKETPINQKWLIDNLD--- 199
 DB 380 IADLNKGDYMS---LTGNIISIINDPS-FKLKDLVE-----QSKRGFLDNLINGVLE 428
 QY 200 ---PDKOTILMSAGATGVSFGFDITDILAKSANAQVVMICGSKELKRSLTAKFKLTR 256
 DB 429 QDIKNSQIIKQQLINYGMEAGDVTKLTKIMP-----ILLDKPESLK-----KVFR 473
 QY 257 MYLILGYTKHNMENWMASSQLMITKPGG-----ITITEGFARCPIMFLNPAP 303
 DB 474 DFTKGNVTK-----MTKELISLTKONPKIKIYLNNNRAIFASILDKTLMDIPGI--NNLD 526
 QY 304 GOELENAFYFEKGFCKIADTPPEAIKIV-----ASLTNGNE-QLTN 344
 DB 527 KQELYN-----LPSMLNHPDELKIVIEVEKSHYHGAVSAIYNLAQNTNFEGLPN 579
 QY 345 MISTMEQDKIKYATQTI-----CRDLDLI 369
 DB 580 IIRAGNSGFNYATEKVKVDVSSSRDFDKV 610
 RESULT 12
 MURG_ENTHR STANDARD; PRT; 360 AA.
 ID MURG_ENTHR
 AC O07670;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

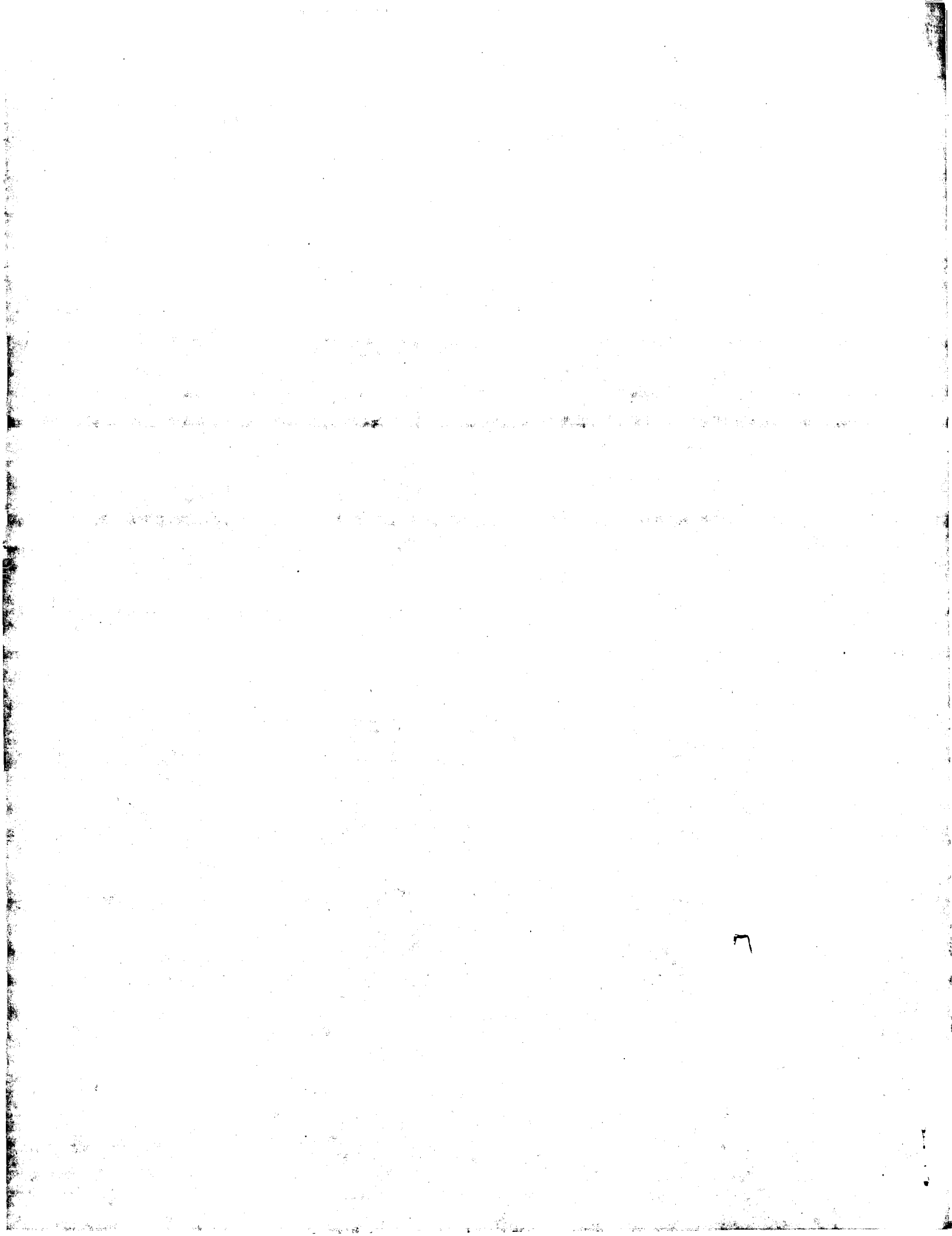
```

30-MAY-2000 (Rel. 39, Last annotation update)
UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGALUCOSAMINE TRANSFERASE
(EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
TRANSFERASE).
MURG.
Enterococcus hirae.
OS
Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC
Enterococcus.
OX
NCBI_TaxID=1354;
ON
[1]
SEQUENCE FROM N.A.
RP
STRAIN=S185;
RC
MEDLINE=99449055; PubMed=10520745;
RX
Duez C., Thamm I., Sapunaric F., Coyette J., Ghuyssen J.-M.;
RT
"The division and cell wall gene cluster of Enterococcus hirae S185.";
RL
DNA Seq. 9:149-161(1998).
CC
-1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
(PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC
-1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC
-1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC
-1- SIMILARITY: BELONGS TO THE MURG FAMILY.
-----
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC
EMBL; Y13922; CAA74235.1;
CR
Transferase; Glycosyltransferase; Cell division; Cell wall;
KW
Membrane; Peptidoglycan synthesis.
SQ
SEQUENCE 360 AA; 39591 MW; 444D6F7CFEC1FDA1 CRC64;

```

MURG_BACSU	STANDARD;	PRT;	363 AA.
ID	MURG_BACSU	STANDARD;	PRT;
AC	P37585; P18578; Q59247;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	UDP-N-ACETYLGALACTOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)		
DE	PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGALACTOSAMINE TRANSFERASE		
DE	(EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC		
DE	TRANSFERASE).		
GN	MURG.		
OS	Bacillus subtilis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=1423;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=92380484; PubMed=1387377;		
RX	Miyao A., Yoshimura A., Sato T., Yamamoto T., Theeragool G.,		
RA	Kobayashi Y.;		
RA	"Sequence of the Bacillus subtilis homolog of the Escherichia coli		
RT	cell-division gene murG.;"		
RL	Gene 118:147-148(1992).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=168;		
RC	MEDLINE=93003529; PubMed=1391053;		
RX	Henriques A.O., de Lencastre H., Piggott P.J.;		
RA	"A Bacillus subtilis morphogene cluster that includes spoVE is		
RT	homologous to the mra region of Escherichia coli.;"		
RL	Biochimie 74:735-748(1992).		
RL	[3]		
RN	SEQUENCE OF 216-363 FROM N.A.		
RP	MEDLINE=90078133; PubMed=2556375;		
RA	Beall B., Lutkenhaus J.;		
RA	"Nucleotide sequence and insertional inactivation of a Bacillus		
RT	subtilis gene that affects cell division, sporulation, and		
RT	temperature sensitivity.;"		
RL	J. Bacteriol. 171:6821-6834(1989).		
CC	-1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC		
CC	INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID		
CC	INTERMEDIATE II) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-PYROPHOSPHORYL-MURNAC-		
CC	(PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II).		
CC	-1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.		
CC	-1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE MURG FAMILY.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; D10602; BAA01454.1; -		
DR	EMBL; X64259; CAA45558.1; -		
DR	EMBL; M31827; AAA83968.1; -		
DR	EMBL; Z99111; CAB13395.1; -		
DR	PIR; JC1275; JC1275.		
DR	PIR; S26499; S26499.		
DR	Subtilist; BG10227; murG.		
KW	Transferase; Glycosyltransferase; Cell division; Cell wall;		
KW	Membrane; Peptidoglycan synthesis.		
FT	CONFLICT 181 181 K -> G (IN REF. 1).		
FT	CONFLICT 270 270 A -> T (IN REF. 1).		
SO	SEQUENCE 363 AA; 39977 MW; 185B20688C732489 CRC64;		

7



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 09:05:02 ; Search time 38.9 seconds
(without alignments)
1329.853 Million cell updates/sec

Title: US-09-668-788-4
Perfect score: 2048
Sequence: 1 MVTQNKILLITGSFGNGHM.....SSQPOEIKYKVLXARFEVK 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL16:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mnc:*
- 8: sp.organella:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.veterebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2048	100.0	391	2	O86492	O86492 staphylococ
2	309	15.1	464	10	Q9S193	Q9S193 arabidopsis
3	309	15.1	455	10	Q9E2L5	Q9E2L5 arabidopsis
4	299	14.6	411	2	Q9RVF3	Q9RVF3 deinococcus
5	296.5	14.5	468	10	O82730	O82730 arabidopsis
6	284	13.9	533	8	Q9MU68	Q9MU68 arabidopsis
7	284	13.9	533	10	O81770	O81770 arabidopsis
8	283	13.8	525	10	P93115	P93115 cucumis sat
9	279.5	13.6	522	10	Q9SM44	Q9SM44 spinacia ol
10	279.5	13.6	530	10	Q9E2L4	Q9E2L4 glycine max
11	278	13.6	535	10	Q9E2L3	Q9E2L3 nicotiana t
12	245	12.0	373	2	O34625	O34625 bacillus su
13	230	11.2	374	2	Q9KBH0	Q9KBH0 bacillus ha
14	194.5	9.5	384	2	Q9EX00	Q9EX00 streptomyce
15	139	6.8	374	2	Q9SS19	Q9SS19 escherichia
16	124	6.1	363	2	Q9K9T0	Q9K9T0 bacillus ha
17	121	5.9	359	2	O87161	O87161 vibrio chol
18	116	5.7	530	6	O18777	O18777 oryctolagus
19	115.5	5.6	398	1	Q58652	Q58652 methanococ

```
001072 kluyveromyc
051410 borrelia bu
09f002 streptococ
006038 lactococcus
09ycs0 aeropyrum p
016917 caenorhabdi
023334 caenorhabdi
09wt34 caenorhabdi
0914h3 brevbacter.
P97886 ratu8 norv
Q45916 clostridium
Q53550 clostridium
Q38197 clostridium
Q92x77 clostridium
018629 caenorhabdi
066802 aquifex aeo
Q9dhg6 yaba-like d
Q93242 caenorhabdi
Q9hfh5 pichia etch
Q69054 human herpe
Q9fdj3 mycoplasma
067183 aquifex aeo
Q9xg35 guillardia
018736 bos taurus
Q9tsm0 macaca fasc
046423 felis silve
```

ALIGNMENTS

RESULT 1

```
O86492
ID O86492 PRELIMINARY; PRT; 391 AA.
AC O86492:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE YPFP PROTEIN.
GN YPFP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=COL;
RC MEDLINE=98313013; PubMed=9650993;
RA Ludovice A.M., Wu S., de Lencastre H.;
RT "Molecular cloning and DNA sequencing of the Staphylococcus aureus
RT UDP-N-acetylmuramyl tripeptide synthetase (mure) gene, essential for
RT the optimal expression of methicillin resistance.";
RL Microb. Drug Res. 4:85-90(1998).
DR EMBL; Y14370; CAA74741.1;
SQ SEQUENCE 391 AA; 44703 MW; 40CDDC37A2627C6C CRC64;
```

Query Match 100.0%; Score 2048; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e-151;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MVTQNKILLITGSFGNGHMQVTSIVNQLNDNLHLSVIEHDLFMEAHPILTICKKW 60
Db 1 MVTQNKILLITGSFGNGHMQVTSIVNQLNDNLHLSVIEHDLFMEAHPILTICKKW 60
QY 61 YINSFKYFRNMKGYYSRPDKLDKCFYKYGLNKLNLILKEKPDILLITFFTPVMSVL 120
Db 61 YINSFKYFRNMKGYYSRPDKLDKCFYKYGLNKLNLILKEKPDILLITFFTPVMSVL 120
QY 121 TEQFNINIPATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPID 180
Db 121 TEQFNINIPATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPID 180
```

QY 181 NKFTPIKQKWLIDNNLDPKQITLMSAGAFVSGKGFDTMTDILAKSANAQVVMICGK 240
DB 181 NKFTPIKQKWLIDNNLDPKQITLMSAGAFVSGKGFDTMTDILAKSANAQVVMICGK 240
QY 241 SKELKRSLTAKFKLTRYMLYLILGYTKHNMENWASSQLMTKPGGTTITTEGFARCPIMIFLN 300
DB 241 SKELKRSLTAKFKLTRYMLYLILGYTKHNMENWASSQLMTKPGGTTITTEGFARCPIMIFLN 300
QY 301 PAPGOELENAFYFEKGGKGTADTPEEAIKIVASLTNGNEQLTNMISTMEODKIKIYATQT 360
DB 301 PAPGOELENAFYFEKGGKGTADTPEEAIKIVASLTNGNEQLTNMISTMEODKIKIYATQT 360
QY 361 ICRDLDDLIGHSSQOEIYGVPLIYARFVK 391
DB 361 ICRDLDDLIGHSSQOEIYGVPLIYARFVK 391
RESULT 2
Q9S193 PRELIMINARY; PRT; 464 AA.
AC Q9S193;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE PUTATIVE MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
GN AT2G11810.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Mea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."
RL Nature 402:761-768(1999).
DR EMBL: AC007187; AAD28678.1; -
DR InterPro: IPR001296; -
DR Pfam: PF00534; Glycos_transf_1;
SQ SEQUENCE 464 AA; 52859 MW; FEC2B424CFBBA136 CRC64;

Query Match 15.1%; Score 309; DB 10; Length 464;
Best Local Similarity 25.4%; Pred. No. 5e-16;
Matches 104; Conservative 83; Mismatches 178; Indels 44; Gaps 16;
QY 6 KKILITGSGFNGHMVQTSIVNQLNDMLDLSVIEHDLFMEAHPIILTSICKKWIYNSF 65
DB 72 KTVLLMSDTGGHRSASAEIRDAFKIEFGDDYRIILKDWKE-----YTGWPLNDM 123
QY 66 -KYFRNMYK-----GFYSRDKLDKCFYK----YVGLNKLINLLIKEKPDILLITFP 113
DB 124 EROYKFMVHVLWSVAFHGTSPKWIHKSYSLSAALYYA-KEIEAGLMYKPDIIISVHP 182
QY 114 ----TPVMSVLTEQFNINIPVATYMTDYR-LHKNMITYSTRYYVATKQDFIDVGD 168
DB 183 LMQHPLWVMKQGLHKKVIFVTITDLNLTCHRTWPHGVSRVCYCPKSEVAKRALVDGLD 242
QY 169 PSTVKVTGIPIDNKP-ETPINQKOWLIDNNLDPKQITLMSAGAFVSGKGFDTMT--DI 225
DB 243 DSQIRVFLGVPVRSFPRTILNKNELKELEIDLNLPAVLLMGEGMGVQKTAALGDS 302
QY 226 L--AKSAN--AQVVMICGSKELKRSITA-KFKLTRYMLYLILGYTKHNMENWASSQLMTK 280
DB 303 LYNKESNPICGLIVICGRNKLVLASTLASHWKIP--VKVRGFETOMKWMGACDCIITK 360

QY 281 PGGITITEGFARCPIMIFLNPAQOELENAFYFEKGGKGTADTPEEAIKIVAS-LTNGN 339
DB 361 AGPCTAAELICGLPIILNDYIPGQKGNVPYVVDNGAGVGTSPKETAIVADWFSNNK 420
QY 340 EQTNMISTMEODKIKIYATQITCRDLDDLIGHSSQOEIYGVPLIYARF 388
DB 421 EEL----KMSNALKLSQPEAVFDIVKDIHLSQQQQ---RIPLNFEF 462
RESULT 3
Q9FZL5 PRELIMINARY; PRT; 465 AA.
AC Q9FZL5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE MGDG SYNTHASE TYPE C.
GN MGDG.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT "The Multigenic Family of MGDG synthases."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047398; BAB12041.1; -
SQ SEQUENCE 465 AA; 52990 MW; 05E0157012E50A14 CRC64;

Query Match 15.1%; Score 309; DB 10; Length 465;
Best Local Similarity 25.4%; Pred. No. 5e-16;
Matches 104; Conservative 83; Mismatches 178; Indels 44; Gaps 16;
QY 6 KKILITGSGFNGHMVQTSIVNQLNDMLDLSVIEHDLFMEAHPIILTSICKKWIYNSF 65
DB 73 KTVLLMSDTGGHRSASAEIRDAFKIEFGDDYRIILKDWKE-----YTGWPLNDM 124
QY 66 -KYFRNMYK-----GFYSRDKLDKCFYK----YVGLNKLINLLIKEKPDILLITFP 113
DB 125 EROYKFMVHVLWSVAFHGTSPKWIHKSYSLSAALYYA-KEIEAGLMYKPDIIISVHP 183
QY 114 ----TPVMSVLTEQFNINIPVATYMTDYR-LHKNMITYSTRYYVATKQDFIDVGD 168
DB 184 LMQHPLWVMKQGLHKKVIFVTITDLNLTCHRTWPHGVSRVCYCPKSEVAKRALVDGLD 243
QY 169 PSTVKVTGIPIDNKP-ETPINQKOWLIDNNLDPKQITLMSAGAFVSGKGFDTMT--DI 225
DB 244 DSQIRVFLGVPVRSFPRTILNKNELKELEIDLNLPAVLLMGEGMGVQKTAALGDS 303
QY 226 L--AKSAN--AQVVMICGSKELKRSITA-KFKLTRYMLYLILGYTKHNMENWASSQLMTK 280
DB 304 LYNKESNPICGLIVICGRNKLVLASTLASHWKIP--VKVRGFETOMKWMGACDCIITK 361
QY 281 PGGITITEGFARCPIMIFLNPAQOELENAFYFEKGGKGTADTPEEAIKIVAS-LTNGN 339
DB 362 AGPGTAAELICGLPIILNDYIPGQKGNVPYVVDNGAGVGTSPKETAIVADWFSNNK 421
QY 340 EQTNMISTMEODKIKIYATQITCRDLDDLIGHSSQOEIYGVPLIYARF 388
DB 422 EEL----KMSNALKLSQPEAVFDIVKDIHLSQQQQ---RIPLNFEF 463

RESULT 4
Q9RVF3 PRELIMINARY; PRT; 411 AA.
AC Q9RVF3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

Qy	6	KKILITGSCFGNHQVQTQSTVQNLQNDMLDLHSVIEHDLFMEAHPILTSICKWYINSF	65
Db	142	KKVILMSDTGGGHRASAEAIRAFNOEFGEYQVFTDLWDHTP-----WPFNOL	193
Qy	66	KYFRN-----MYKGYSRPOKL-----DKCFYKYGLKNLLIKLKEPKDLILITFP-	113
Db	194	PRSYNELVKHCTLWKMTYYGTSPIRVHOSNFAATSFIAEIAQGLMKQPDIIISVHPL	253
Qy	114	---TPVMSVLTEQNNINIPATVMTDK-RLHKNNIIPYSYRYVATKETKQDFIDVGDP	169
Db	254	MOHVPLRLVRSKGLLKIVFTTITDLSTCHPTWFHKLTRCYCPSTEVAKRAQAGLET	313
Qy	170	STVAVTGIPIDNKFETPINOKWL-----IDNNLDPDKOTILMSAGFV-----SKGF	218
Db	314	SOIKYGLPVRPSEVKRVLRELGMENL----PAVILMGGEQGMGPIETARAL	369
Qy	219	DTMITDLAKSANAQVVMICGSKELKRLSLTA-KFKLTRMYILGYTKHKHNEWMASQLM	277
Db	370	ADALYDKNLGEAVQVLIICGRNKKLQSLSSLDWKIP--VQVKGFITKMEECMGACDCI	427
Qy	278	ITKPGGTTITEGFARCIPMFLNAPQOELENAFYEFGFKIADTDPPEAKIVASLTN	337
Db	428	ITKAGPOTIAEAMIRGLPIILNGYIAQGEAGNVPYVYVENGCGKFSKREISKIVADWFG	487
Qy	338	GNEQLTNMISTMEQDKIKYATQ----TICRLDLDLI-GHSSOPQ	376
Db	488	---PASKELEIMSQNALRLAKPEAVFKIVDHMVELYRKKNSLPQ	528

RESULT 7

O81770

ID

O81770

PRELIMINARY;

PRT;

533 AA.

AC

DB

O81770;

DT

01-NOV-1998

(T-EMBLrel. 08, Created)

DT

01-NOV-1998

(T-EMBLrel. 08, Last sequence update)

DT

01-MAR-2001

(T-EMBLrel. 16, Last annotation update)

DE

MONOGALACTOSYDIACYLGLYCEROL SYNTHASE - LIKE PROTEIN

(MONOGALACTOSYDIACYLGLYCEROL SYNTHASE-LIKE PROTEIN) (MGDG SYNTHASE TYPE A).

DE

EN

F28M20.30 OR AT4G31780 OR MGDA.

GN

Arabidopsis thaliana (Mouse-ear cross).

OC

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsides.

OX

NCBI_Taxid=3702;

ON

[1]

RP

SEQUENCE FROM N.A.

RP

Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Hoheisel J., Mewes H.W., Mayer K.F.X., Schueller C., Bevan M.;

RA

Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.

RN

[2]

RP

SEQUENCE FROM N.A.

RP

Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W., Lemcke K., Mayer K.F.X.;

RA

Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RN

[3]

RP

SEQUENCE FROM N.A.

RP

Terryn N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R., Glielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;

RA

Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RN

[4]

RP

SEQUENCE FROM N.A.

RA

EU Arabidopsis sequencing project;

RL

Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.

RN

[5]

RP

SEQUENCE FROM N.A.

RP

Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;

RT

"The Multigenic Family of MGDG synthases".

RL

Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

DR

EMBL; AL031004; CAA19745.1; -.

DR

EMBL; AL161579; CAB79896.1; -.

DR	EMBL; AB047399; BAB12042.1.; -;
DR	Mendel; 32401; Arath;1856;32401.
SQ	SEQUENCE 533 AA; 58537 MW; E581E67317CB9CC8 CRG64;
	Query Match 13.9%; Score 284; DB 10; Length 533;
	Best Local Similarity 25.5%; Pred. No. 5.4e-14;
	Matches 103; Conservative 66; Mismatches 185; Indels 50; Gaps
Qy	6 KKIIITIGSPGNHGMYQTQSIVNQLNDMNDLHLSVIEDHLFMEAHPIILTSICKKWYNSF 65
Db	142 KKVLLMSDGTGGHRSASAEAIRAFNQEFGEVQVFITDLWTHTP-----WPFNL 193
Qy	66 KYFRN-----MYKGYYSRDPKL-----DKCFKYKYGLNKLNLLTIKEKPDILLTFP- 193
Db	194 PRSYNFVLKHGTLLWKMTYYTSPRIVHQSFNAATSTFIAREIAOGLMKYPDIISVHPL 253
Qy	114 ---TPVMSVLTEQFNINIPATVMTDY-RLEKKNIIPYSTRYVVAVKETKDQFDVGIDP 169
Db	254 MQHVPLRVLSKGLLKIVTTTIDLSTCPHFHKLVTRCYPCSTEVAKRAQAAGLSET 313
Qy	170 STVKVTGIPIDNKPETPINOKWL-----IDNNLDPKDTILMSAGFY-----SKGF 218
Db	314 SOIKVYGLPVRSFPVKPRKVELRLGLMDENL----PAVLLMGEGEGMPIEARAL 369
Qy	219 DTMITDILASANAQVMICGCKSELKRSLTA-KFKLTRMWYLILGYTKHNEMWASSQLM 277
Db	370 ADALYKDNLGEAVGVLIICGRNKKLQSKLUSSLDWKIP--VQVKGITKMEECMGACDCI 427
Qy	278 ITRPGGTTITEGARCIPTMFLPAPQOELENAYFEPEEKFGFGKIADTPEAIVKIVASLTN 337
Db	428 ITRAGPCTIAEAMIRGLPIILNGYIAGQEAAGNVPPYVENCGCKFSKSREISKIVADWF 487
Qy	338 GNEOLTNWTSMQDKKIKYATQ-----TICRDLDLI-GHSQQPQ 376
Db	488 ---PASKEFLMSQNARLAKPEAVFKIVDHMHVELVRKKNLSLPQ 528
RESULT 8	
P93115	
ID	P93115 PRELIMINARY; PRT; 525 AA.
AC	P93115;
CD	01-JUN-1998 (TREMBLrel. 06, Created)
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE	MONOGALACTOSYDIACYLGlycerol SYNTHASE PRECURSOR (EC 2.4.1.46)
DE	(1,2-DIACYLGlycerol 3-BETA-GALACTOSYLTTRANSFERASE).
OS	Cucumis sativus (Cucumber).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OX	Cucurbitales; Cucurbitaceae; Cucumis.
OX	NCBI_TaxId=3659;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	STRAIN=RONGAJIBAI;
RX	MEDLINE=97144442; PubMed=6990209;
RA	Shimojima M., Ohta H., Iwamatsu A., Masuda T., Shioi Y.,
RA	Takamiya K.-I.;
RT	"Cloning of the gene for monogalactosyldiacylglycerol synthase and its
RT	evolutionary origin.";
RL	Proc. Natl. Acad. Sci. U.S.A. 94:333-337(1997).
CC	-!- FUNCTION: CATALYZES THE FORMATION OF MONOGALACTOSYLDIACYLGlycerol
CC	(MGDG) WHICH IS A MAJOR STRUCTURAL LIPID OF THE CHLOROPLAST.
CC	-!- CATALYTIC ACTIVITY: UDP-GALACTOSE + 1,2-DIACYLGlycerol = UDP + 3-
CC	BETA-D-GALACTOSYL-1,2-DIACYLGlycerol.
CC	-!- PATHWAY: GLYCOSYLATION.
CC	-!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC	-!- SIMILARITY: TO E.COLI AND B.SUBTILIS UDP-N-ACETYLGlucoSAMINE--N-
CC	ACEYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-
CC	ACEYLGlucoSAMINE TRANSFERASE.
DR	EMBL; U62622; AAC49624.1; -;
DR	Mendel; 12656; Cucsa;1856;12656.
KW	Transferase; Glycosyltransferase; Transit peptide; Chloroplast

Db 371 YDENIGAPVQIILVICGRNKKLANKLSSINWKP--VQVGFVTKMECGACDCIITKA 428
QY 282 GGTITTEGFARCPIMFLNPAPQOELENAFYFEKFGKIADTPBEAIVASLTNGNEQ 341
Db 429 GPGTIAEQIRGLPIILNDYIAGOEAGNVPVYVYENGCGFKSPKDIKIAVAFWFGPK--486
QY 342 LTNMISTMEQDKIKYATQ-----TICRDLDDLIGHSS 373
Db 487 -AYELQOMSONLARLPDPAVFKIVHDLHELVRQRS 521
RESULT 11
Q9FZL3
ID Q9FZL3 PRELIMINARY; PRT; 535 AA.
AC Q9FZL3
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
DE MGDG SYNTHASE TYPE A.
GN NTMGD A.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
OX NCB_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai K., Takamiya K., Ohta H.;
RT "cDNA cloning of MGDG synthase from tobacco and soybean.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047476; BAB11980.1; -;
SQ SEQUENCE 535 AA; 59589 MW; B1B2067E86EDE477 CRC64;
Query Match 13.6%; Score 278; DB 10; Length 535;
Best Local Similarity 23.5%; Pred. No. 1.6e-13;
Matches 98; Conservative 78; Mismatches 182; Indels 42; Gaps 13;
QY 6 KKLIIITG-SFGNGHMQVTSIVNQLMDLHLSVIEHDLFMEAHPIILTSICKKRYIN-- 63
Db 145 KKVILMSDTGGHRSASAEIRSAFNEFGDKYQVFIIDLWTEHP-----WPFNQL 196
QY 64 --SKFY--RNMVGFYSRPDKL-----DKCFYKYVGLINKLINLIIKEKPLIILTP- 113
Db 197 PHSYFLVKHGLWRTYIATAPRLVHQTNFAATSTFIAREVAKGLMKYQPDIIISVHPL 256
QY 114 ----TPVMSVLTQFNINIPVATMTDY-RLHKNWITPYSTRYVATKTKQDFIDVGIDP 169
Db 257 MQHVPLRLRSKGLLKIIFTVTITDLSCTHPTWFKLVTRCYCPSEEVAKRALRAGLKP 316
QY 170 STVKYTGIPIDNKFETPINQKWL-IDNNLDPDKQTIILMSAGAFV-----SKGFDIMI 222
Db 317 YOLKVTGLVPSFVKVPVPPKVELRKEGLMEHPLPAVILMGSGGGMPIEATARALGDAL 376
QY 223 TDLAKSANAQVVMICGSKSKELRSITA-KFKLTRMYLILGYTKHNMENWMASSQLMINKP 281
Db 377 YDEIHGEPIQVLVICGRNKKLFNLTSTVQWIP--VQVGFVTKMECGACDCIITKA 434
QY 282 GGTITTEGFARCPIMFLNPAPQOELENAFYFEKFGKIADTPBEAIVASLTNGNEQ 341
Db 435 GPGTIAEQIRGLPIILNDYIAGOEAGNVPVYVYENGCGFKSPKDIKIAVAFWFGPRQ- 493
QY 342 LTNMISTMEQDKIKYATQ-----TICRDLDDLIGHSS-OPQ 376
Db 494 --DELIRMSONLARLPDPAVFKIVHDMHELVRORNFEPQ 531
RESULT 12
O34625
ID O34625 PRELIMINARY; PRT; 373 AA.
AC O34625
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)

DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
DE YKON.
GN YKON.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCB_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blandhard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapina A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Prescann E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconelli E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ002571; CAA05612.1; -;
DR EMBL: Z99111; CAB13208.1; -;
DR EMBL: Z99110; CAB13192.1; -;
SQ SEQUENCE 373 AA; 43480 MW; 62FFCF2BDC598CDC CRC64;

Query Match 12.0%; Score 245; DB 2; Length 373;
Best Local Similarity 23.5%; Pred. No. 3.5e-11;
Matches 88; Conservative 75; Mismatches 168; Indels 44; Gaps 14;
QY 6 KKLIIITG-SFGNGHMQVTSIVNQLMDLHLSVIEHDLFMEAHPIILTSICKKRYIN 64
Db 2 KNLIFPFLSTSTGHHVADALQAELESQ--LAAEKIDIFSHSYRLEKLSVALKW 58
QY 65 FKYPNNMYKGFYSRDKL-----DKCFYKYIGL--NKLINLLKEKPDILLT--F 112
Db 59 IOYFPKTYSGIY-----RLACGEFQHDKRYFYECFTQOMRHILQEKQPDIAFCFHAL 113
QY 113 PTPVMSVLTQFNINIPVATMTDYRLHKNWITPYSTRYVATKTKQDFIDVGIDPSTV 172
Db 114 PSYLLNRLKEYP-NTLVNVNVTDFVNLWGRKNIDYHFVPSTEVRKQLISEGIDONNI 172

```

RESULT 1/4
Q9EX00 PRELIMINARY; PRT; 384 AA.
ID Q9EX00
AC Q9EX00;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
OS 2SCG38.19C;
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces
OX NCBI_TaxID=1902;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RA Saunders D.C.; Harris D.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RA Cerdeno A.M.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M.; Kieser H.M.; Denapaite D.; Eichner A.; Cullum J.;
RA Kinashi H.; Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL445503; CAC13078.1; -.
SQ SEQUENCE 384 AA; 40689 MW; E64E01E05B1DDE40 CRC64;
Query Match 9.5%; Score 194.5; DB 2; Length 384;
Best Local Similarity 20.2%; Prd. No. 3.1e-07;
Matches 72; Conservative 73; Mismatches 160; Indels 51;

```

Qy	6	KKILLITGFCNGHMQVTSVINQLND-----MNLDSLVSIEHDLFMEAHPIILTSICKW	60
Dd	15	RRLIVISASMGAGHD7VAALVRRAREGD7TAQTVDVLALLPYG-----LGAVLRCF	66
Qy	61	YINSFKYERNMYKGFY--YSRDPDKLDCFKYYCYL-----NKLINLLIKEK	104
Dd	67	YRGSVRIHPWAYAALYLRLFLRP-----GAGRRPSTPLARLAGDRLELAARTG	115
Qy	105	PDLILLTF--PTPYMSVLTEQFINIPVATVMTDYIRLHKNNWITPYSTRYYVATKETKQDF	162
Dd	116	ADVVPVPHLGAQLTHLRDGRLLPVPSVVLVIDFELHQWLHPGNDHCLCTEEAAAREA	175
Qy	163	I-DVGIDPSVTKVTGIPIDNKFTETPINQOKWL-IDNNLPDPKOTILMSAGAFVSKGFDI	220
Dd	176	RNGTGTAEZCGPVVAPEFSAGRVP-GAAOWRETFDLGGPRFAVVLISAGAWGVSHLDA	234
Qy	221	MITDILAKSANAQVMVMTCGSKEKLKRSLTRAKFILTRYLIJGYTKHNEMWASSQLMITK	280
Dd	235	TVR-LLVDHGYPVLVLCGDNQRLLRTLSG----TPGVLAGWVTDMPGLHAARALIDN	288
Qy	281	PGGITTEGARCIPIFMFLAPAQOELENAFYEFGKGFIADTPEEAIKIVASLT	336
Dd	289	AAGTAVQAALAAGLPVVGHRPIPGHGADGVRRMAALGVSEVADRTELLEALART	344
RESULT	15		
Q9S519	ID	PRELIMINARY;	PRT; 374 AA.
AC	Q9S519;		
DT	01-MAY-2000	(TEMBLrel. 13, Created)	
DT	01-MAY-2000:	(TEMBLrel. 13, Last sequence update)	
DE	01-MAR-2001	(TEMBLrel. 16, Last annotation update)	
DE	PUTATIVE GLYCOSYL TRANSFERASE.		

```
GN WBDM.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STOKE W, M92;
RX MEDLINE=96060831; PubMed=7590310;
RA Bastin D.A., Reeves P.R.;
RT "Sequence and analysis of the O antigen gene (rfb) cluster of
RL Escherichia coli O111.";
RL Gene 164:17-23(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-STOKE W, M92;
RX MEDLINE=98449835; PubMed=9774562;
RA Wang L., Curd H., Qu W., Reeves P.R.;
RT "Sequencing of Escherichia coli O111 O-antigen gene cluster and
RL Identification of O111-specific genes.";
RL J. Clin. Microbiol. 36:3182-3187(1998).
DR EMBL: AF078736; AAD46732.1; -.
DR InterPro: IPR001296; -.
DR Pfam: PF00534; Glycos_transf_1; 1.
KW Transferase.
SQ SEQUENCE 374 AA; 42192 MW; 54C0C1EFBBF76B53 CRC64;

Query Match 6.88; Score 139; DB 2; Length 374;
Best Local Similarity 20.9%; Pred. No. 0.0062;
Matches 84; Conservative 69; Mismatches 163; Indels 86; Gaps 18;

QY 6 KKILITGFCNGHMOVTQSTIVNOL-----NDMLNDHLSVIEHDLFMEAHPLT/SICKKWY 61
DB 2 KIVYIIIGLTCGGAHELMFTQLADQMFIRGHVDVNIICLTGIS-----EVKP--TQNIHY 54
QY 62 INSKFYRNMYKGYYSRDPDKLDCFKYKYLKLNLLIKERPDLI-----L 109
DB 55 VMKDKNFRSFRALF-----QVKKIIVALKPDIHSHMFHANIESRF 96
QY 110 LTFFPTPVMSVLTEQFNINIPVATVMTDYRLHKWITPYSTRYYVATRETQDFIDVGIDP 169
DB 97 IRMLIPAVPLICTAHNKEGNGARMFCYRL--SDFLASITN---VSREAVQEFYARKATP 152
QY 170 STVKVTGIP---IDNKFTPINQKWLIDNNLDPDKQITILMSAGAFGVSKGFDITMTD-- 224
DB 153 KN-KIVEIPNFINTKPFDINVRKKTDAFNLDSTAVLLAVGLVEAKDYPNLLNAIN 211
QY 225 --ILAKSNAQ--VVMICG----KSKELKRSLTAKFKLTRMYLILGYTKHNEWMASQL 276
DB 212 HLILSKTSNCNDFILLIAGDGALRNKLL--DLVQQLNLVDKVFVFGORSIDIKELMCAADL 269
QY 277 MTRPG----GITITEGFARCIPIFLNPAQGE-----LENAFYEEKGFCKIA 322
DB 270 FVLSEWEGFGLVVAEAMACERPVVATDSGGVKEVGPVPHNDVIPVNSHILLAE----KIA 325
QY 323 DT---PEAAIKIVASLTN---GNEQLTNMISTMEQDKIKYA 357
DB 326 ETLKIDDNARKIIGNKREIYVNSFINSIVSEWERLYFKYS 367
```

Search completed: June 29, 2001, 09:05:04
Job time: 321 sec


```

QY      141 KNNITPYSTRYYVATK--ETKQDFIDV-----GIDPSTVKVTGI-- 177
      : ||| : :| : :| : :| : :| : :| : :| : :| : :| : :|
Db      223 LS-ITPYES---LASELLQREMSLVEVLSHASVWLFGRGDFVDFYPRPIMPVFI GGNC 278
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
QY      178 ----PIDNKPETIPNQKWLIDNNLDPDKOTILMSAGAFGV---SKGFDTMITDILAKSA 230
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
Db      279 VIKKPLSQEEFAYVN-----ASGEHGIVVFSLG--SWVSEIPEKKA 317
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
QY      231 NAQVWICKSKSKELKSLTAKFKLTRMYLILGYTKHNMENWMASSQLM-----ITKPG 282
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
Db      318 -MEIABALGR--IPOTILWRVTGTSPSLAKNT-ILVKLPQNDLLGHKPAFTIHSG 372
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
QY      283 GIITGEFARCIPIMLFNPAQGLBNAFVFEKGGF-----KTADPPEAIKIVASL 335
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
Db      373 SHGIYEGICNGVPMVM-PLFGDQMDNAXMETRGAGVTLNVLEMTADDLENALKTVINN 431
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
QY      336 TNGNEQLTNMISTMEQDK 353
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|
Db      432 KSYKENIMR-LSSLHKDR 448
      : :| : :| : :| : :| : :| : :| : :| : :| : :| : :|

RESULT      4
PCT-US92-00282-7
; Sequence 7, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UCT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-00282-7

```

QY	64	-----SFKYFRNMYKGF-----YYSRPDKLDKCFYKYYGLNK-----L 96
Db	75	KSPFPVYNLSELETRYRSKFSGFNHFAASSPLMAPLREYRNNMIVIDMCFSCQSLKDSAT 134
QY	97	INLLIKPKDILLTTPPTVMSVLTBQFNI-----NIP--- 129
Db	135	LSFLRNOQFALTDPAFCGVLAELYKLPSIYLFGRGFCPSLEHIGQSPSPVSVYPRPY 194
QY	130	-----VATVMTDYLHKHNMTPYSTRYYVATKTKQDFIDVGDIPSTVK 173
Db	195	TKFSDHMTFPQRLANTIANILENYLH---CLYSKYEILASDLLKRDVSLPALHONSLLW 250
QY	174	VTGIPIDNKFETPINQKOWLI-----DNNLDPPDKQTILMSAGAFV---SKGFDMTIT 223
Db	251	LLRYDFVFEYPRPWPNNMIFIGTGNCKKGNLSQEFEEAVVNASGEHGVVVSIG--SMVS 308
QY	224	DILAKSANAQVVMICGSKSLKSLTAKFLKTRMYLILGYTKHMNENWMASSQLM----- 277
Db	309	EIPEKKA-MEIAEALGR--IPQTLWRVYTGTRPSNLAKNT-ILVKWL PQNDLLGHPKAR 363
QY	278	--ITKGGITITGFCARCPIPMIFELNAPQOELENAFYFEKGF-----KIADTPEEA 328
Db	364	AFITHSGSHGIYEGICNGVPMVM--PLFGQMDNAKRMETRGAGVTLNVLMTAADDLENA 422
QY	329	IKIVASLTNGNEQLTNNMISTMEQDK 353
Db	423	LKVTVNNKSYKENIMR-LSLHKDR 446
RESULT 5		
US-08-457-176-2		
; Sequence 2, Application US/08457176		
; Patent No. 5591826		
; GENERAL INFORMATION:		
; APPLICANT: Vogelstein, Bert		
; APPLICANT: Kinzler, Kenneth W.		
; APPLICANT: de la Chappelle, Albert		
; TITLE OF INVENTION: Mutator Gene and Hereditary		
; TITLE OF INVENTION: No. 5591826-Polypsis Colorectal Cancer		
; NUMBER OF SEQUENCES: 16		
; CORRESPONDENCE ADDRESS:		
; ADDRESSEE: Banner, Birch, McKie, and Beckett		
; STREET: 1001 G Street, N.W.		
; CITY: Washington		
; STATE: D.C.		
; COUNTRY: USA		
; ZIP: 20601		
; COMPUTER READABLE FORM:		
; MEDIUM TYPE: Floppy disk		
; COMPUTER: IBM PC compatible		
; OPERATING SYSTEM: PC-DOS/MS-DOS		
; SOFTWARE: PatentIn Release #1.0, Version #1.25		
; CURRENT APPLICATION DATA:		
; APPLICATION NUMBER: US/08/457,176		
; FILING DATE: 01-JUN-1995		
; CLASSIFICATION: 530		
; PRIOR APPLICATION DATA:		
; APPLICATION NUMBER: US 08/160295		
; FILING DATE: 02-DEC-1993		
; ATTORNEY/AGENT INFORMATION:		
; NAME: Kagan, Sarah A.		
; REGISTRATION NUMBER: 32,141		
; REFERENCE/DOCKET NUMBER: 01107.44900		
; TELECOMMUNICATION INFORMATION:		
; TELEPHONE: 202.508.9100		
; TELEFAX: 202.508.9299		
; TELEX: 197430 BBMB UT		
; INFORMATION FOR SEQ ID NO: 2:		
; SEQUENCE CHARACTERISTICS:		
; LENGTH: 934 amino acids		
; TYPE: amino acid		
; STRANDEDNESS: single		
; TOPOLOGY: linear		

; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-457-176-2

Query Match 5.0%; Score 102.5; DB 1; Length 934;
Best Local Similarity 22.3%; Pred. No. 0.15; Indels 93; Gaps 17;
Matches 77; Conservative 45; Mismatches 131; Indels 93; Gaps 17;

QY 18 GHMQVTSIVNQLNDMNL--DHLNVIE-----HDLFMEAHPIITSTCKKWI 62
DB 338 GORLVNQWIKQPLMDKNRIERLNLVEAFVDEALRQTLQEDL-LRRFPDLNRLAKK--- 393
QY 63 NSFYFRNMYKGFYSPDKLDCFKFYKYGKLNKLNLL-IKEKPD-----LILTFPTPV 116
DB 394 -----FORQANLQDCYRLYQGINLPNVIQALEKHEGKHQKLLAVFVTP 440
QY 117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYVATKQDFIDVGIDPSTVKV 174
DB 441 TDLRSDFSKFQEMETTLDMQOVENHEFLVKP-----SFPNLSL 481
QY 175 TGI--PIDNKFE--TPINOKOWLIDNNDPKQTLMSAGAFVSKGFDITMTDILAKSAN 231
DB 482 REIMNDLEKKMQSTLISAAR--DIGLDPGKQIKLDSSAQFGYFRVTCREKVLNKN 538
QY 232 AQVVMICGSKELKRLSTAKFKLRMYLILGYTKHMEWMASSOLMTKPGIITGFA 291
DB 539 FSTVDIQKNGVKFTNS-----KLTS--LNEEYTKNTEYEAAQDAIVKEI--VNISGYV 589
QY 292 RCIPMIFLNPAGOLENAFYFE-----EKGFGKI 321
DB 590 E--PMQTLNDVLAQ-LDAVSVFAHVSNGAPVYVYRPAILEKGQGR 632

RESULT 6

US-08-457-175-2
; Sequence 2, Application US/08457175
; Patent No. 5693470
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: Mutator Gene and Hereditary
; TITLE OF INVENTION: No. 5693470-Polyposis Colorectal Cancer
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie, and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,175
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160295
; FILING DATE: 02-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.44900
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100

; TELEFAX: 202.508.9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 934 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-457-175-2

Query Match 5.0%; Score 102.5; DB 1; Length 934;
Best Local Similarity 22.3%; Pred. No. 0.15; Indels 93; Gaps 17;
Matches 77; Conservative 45; Mismatches 131; Indels 93; Gaps 17;

QY 18 GHMQVTSIVNQLNDMNL--DHLNVIE-----HDLFMEAHPIITSTCKKWI 62
DB 338 GORLVNQWIKQPLMDKNRIERLNLVEAFVDEALRQTLQEDL-LRRFPDLNRLAKK--- 393
QY 63 NSFYFRNMYKGFYSPDKLDCFKFYKYGKLNKLNLL-IKEKPD-----LILTFPTPV 116
DB 394 -----FORQANLQDCYRLYQGINLPNVIQALEKHEGKHQKLLAVFVTP 440
QY 117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYVATKQDFIDVGIDPSTVKV 174
DB 441 TDLRSDFSKFQEMETTLDMQOVENHEFLVKP-----SFPNLSL 481
QY 175 TGI--PIDNKFE--TPINOKOWLIDNNDPKQTLMSAGAFVSKGFDITMTDILAKSAN 231
DB 482 REIMNDLEKKMQSTLISAAR--DIGLDPGKQIKLDSSAQFGYFRVTCREKVLNKN 538
QY 232 AQVVMICGSKELKRLSTAKFKLRMYLILGYTKHMEWMASSOLMTKPGIITGFA 291
DB 539 FSTVDIQKNGVKFTNS-----KLTS--LNEEYTKNTEYEAAQDAIVKEI--VNISGYV 589
QY 292 RCIPMIFLNPAGOLENAFYFE-----EKGFGKI 321
DB 590 E--PMQTLNDVLAQ-LDAVSVFAHVSNGAPVYVYRPAILEKGQGR 632

RESULT 7

US-08-709-784-1
; Sequence 1, Application US/08709784
; Patent No. 6048701
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University
; TITLE OF INVENTION: Antibody Detection of Mismatch Repair
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,784
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,351
; FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.57434
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-709-784-1

Query Match 5.0%; Score 102.5; DB 3; Length 934;
Best Local Similarity 22.3%; Pred. No. 0.15;
Matches 77; Conservative 45; Mismatches 131; Indels 93; Gaps 17;
QY 18 GNCHQVTSIVNQLNDMNL--DHLNVIE-----HDLFMEAHPIILTSICKKWI 62
DB 338 GQRLVNWQIKQPLMDKNRIERLNLVEAFVDAELRQTLQEDL-LRRFPDLNRLAKK--- 393
QY 63 NSFKYRNMYKGFYSPDKLCKFYKYYGLNKLINLL-IKKPD-----LILLTPTPV 116
DB 394 -----FORQAAANLQDCYLYQGINOLPNVIOALEKHEGKQKLLIAVFTPL 440
QY 117 MSVLTE--OFNINIPVATVMTDYRLHKNWITPVSTRYVATKTKODFDVGDIDPSTVKV 174
DB 441 TDLRSDFSEFQEMETTLDMQDVENHEFLVKP-----SFDPNLSEL 481
QY 175 TGI--PIDNKE-TPINQKOWLNDNPDKQITLMSAGAFVSGKGFDTMTDILAKSAN 231
DB 482 REIMDLKKMSTLISAAR--DLGLDPGKQIKLDSSAQFGYFRVTCCKEKLRLNNKN 538
QY 232 AQVVMICGSKSLKSLTAKFKLTRMYLILGYTKHNEHMASSOLMTKPGTITTEGA 291
DB 539 FSTVDLQKGVKFTNS-----KLTS--LNEEYTKNTEYEAAQDAIVKEI--VNISSGYV 589
QY 292 RCIPMIFLNPAPQCELENAFYFE-----EKGFQKI 321
DB 590 E--PMOTLNDVLAQ-LDAVVVSAHVSNGAPVPVVRPAILEKGQRI 632

RESULT 8
US-08-846-762-85
Sequence 85, Application US/08846762A
Patent No. 5994072
GENERAL INFORMATION:
APPLICANT: Lam, Joseph S.
APPLICANT: Burrows, Lori
APPLICANT: Charter, Deborah
TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
FILE REFERENCE: 6580-089
CURRENT APPLICATION NUMBER: US/08/846,762A
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 85
LENGTH: 376
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-08-846-762-85

Query Match 5.0%; Score 101.5; DB 2; Length 376;
Best Local Similarity 22.4%; Pred. No. 0.047;
Matches 53; Conservative 42; Mismatches 79; Indels 63; Gaps 12;
QY 93 LNKLINLLIKERPDILL--TFPTVMVSLTEQFNINIPVATVMTDYR---LHKNW----- 143
DB 75 LEGLKFIKLAEPDVLVHGDDTTTLATSLAAFYQRIQIPGVHVEAGLRTGDLXSPWPEAN 134
QY 144 --ITPYSTRYYVATKET-KODFIDVGDIDPSTVKVGTIPIDNKFPTPINKQWLIDN---- 196
DB 135 RLTGLHLMYHFSPTETSRQNLLENVADSRIFITG-----NTVIDALLWVRDQVMS 187
QY 197 -----NLDPDKOTILMSA---GAFGVSKGFDTM---ITDILAKSANAQVVMIC 238
DB 188 DKLRSELAANYPFIDDPKKMLVTHRRRESFG--RGFEICHALADIATHODIQIVPV 245
QY 239 GSKSEIKRSLTAKFKLTRMYLILGYTKHM-----NEW-----MASSOLMTKPGI 284
DB 246 HLPNVNREPVR-----ILGHVKNVILIDPQYLPFVWLMNHAWLILTDSGGI 293

RESULT 9
PCT-US92-00282-3
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 533 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-3

Query Match 4.9%; Score 99.5; DB 5; Length 533;
Best Local Similarity 19.3%; Pred. No. 0.13;
Matches 74; Conservative 74; Mismatches 132; Indels 103; Gaps 20;
QY 16 GNCHQVTSIVNQLNDMNLHLSVIEHDLFMEAHPIILTSICKKWIINSFKYRNMYKGF 75
DB 126 GCSHLLHNLKELMASLAESSFD--VMLTDPFLPCSPVIAQ-----YLSLPTVF-----F 171
QY 76 YYSRDPKLD----KC--FYKYYG-----LNKLINLLIKEPDLILLTPTTPVM 117

Db 172 LHALPCSLFEATQCPNPFVSVPRLSSSHSDMTFLQVRKNMLIAFQSNFLCDWVSPYA 231
Qy 118 SVLTQENINIPVATVMDRLHKNWTPYSTRYYVATKTKQDFI---DVGIDPSTVKV 174
Db 232 TLASEFLQRETVODLLSASV---WL-----FRSDFVKDYPRPIMPNNMVF 275
Qy 175 TGI-----PIDNKFETPINQKOWLIDNLDPPDKQITILMSAGAFGV---SKGFDTMITDI 225
Db 276 GGINCLHONPLSQEFAYIN-----ASGEHGIIVFSLG--SWVSEI 314
Qy 226 LAKSANAQVVMICGSKELKSLTAKEFLTRMYLILGYTKHMEWMASSQLM----- 277
Db 315 PEKKAMA-IADALCKNQ---TVLWRYTGTTPSNLANNT-ILVKWLPQNDLIGHPMTRAF 369
Qy 278 ITKPGGTYITBGFARCPIMFLNPAPQOELNAPYFEKGFQ-----KIADTPERAIK 330
Db 370 ITHAGSHGVESICNGVPMVM--PLFGQMDNKRMTKGAGVTLNVLMTSEDLENALK 428
Qy 331 IVASITNGNEOLTNMISTMEODK 353
Db 429 AVINDKSYKENIMR-LSSLHKDR 450

RESULT 10
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHAY, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480, 604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.

; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-10

Query Match 4.8%; Score 98.5; DB 1; Length 2366;
Best Local Similarity 20.0%; Pred. No. 1.7; Indels 155; Gaps 24;
Matches 89; Conservative 73; Mismatches 129; Indels 155; Gaps 24;

Qy 2 VTQNKKIL-----IITGSGFGHMQVTSIVNQLNDMLDLHLSVIEHDLFMEAHPI 52
Db 1377 IEEKILNSHEINFSGEVNGS--NGFVSLTFSILEGIN-----AIEVDLLSKSYKL 1427
Qy 53 LTS-----TCKWYINSFYFRMYKGYFYSRDPDKLDCFKYKYGLNKLIN-- 98
Db 1428 LISGELKILMLNSNHIQOK--IDYIGFNSLQKNIPYSFYDSECK-----ENGFINGS 1478
Qy 99 ---LLIKEKPDILLITFTPTVMKSVLTEQ-----FNINIPVATVMTDYRLHKNWITPY 147
Db 1479 TKGLFVSELPDVLIS-----KYMDSDKPSFGYYSNNLKDVKVITKDNVN-----I 1526
Qy 148 STRYYVATKETQDFIDVGI-----DPSTVKVTGIPIDNKFETPINQKOWLIDNLDPPDK 202
Db 1527 LTGYL-----KDDIKISLSLTLDQDEKTIKLSNVHLD---ESGVAR----- 1564
Qy 203 QTILMSAGAFVSGKGTMTDILAKSANAQVVMICGSKELKSLTAKEFLT----- 255
Db 1565 --ILKFMNRKGTNTSDSLMS--FLESNNIKSIFVFNLOSNIRKIFLDANFIISGTTISIGQ 1620
Qy 256 -----RMYLILGYTKHME--WMASSQLMITKP-----GGITIT----- 287
Db 1621 FETICDENDNIQPYFIKFNTELTNYLYVGNRQNMIVEPNVDLDDSGDISSTVINFQKY 1680
Qy 288 -EGFARCPIMFLNP-----APQOELNAP-----YFEKGFKTKADTPEEA 328
Db 1681 LYGIDSCVNRKVIWSPNIYTDENITPYETNTYPEVILVDANYINEKINVNINDL---S 1737
Qy 329 IKIVASLTNGNEOLTNMISTMEODKI 354
Db 1738 IRVWVS-NDGNDFI--LWSTSEENKV 1760

RESULT 11
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 42A
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-10

Query Match 4.8%; Score 98.5; DB 2; Length 2366;

Best Local Similarity 20.0%; Pred. No. 1.7;
Matches 89; Conservative 73; Mismatches 129; Indels 155; Gaps 24;

QY 2 VTQKKIL-----IITSGNGHMQVTSIVNQLNDMLDHLVSIEHDLFMEAHPI 52
Db 1377 IENKIILNSHINFSGEVNGS--NGFVSLTFSILEGIN-----AIEVDLLSKSYKL 1427
QY 53 LRS-----ICKKWIINSFYFRNMYKGFYISRPDKLDKCFYKYGLNKLIN-- 98
Db 1428 LISGELKILMLNSNHIQOK--IDYIGFNGELQKNIPYSFVDSGK-----ENGFINGS 1478
QY 99 ----LLIKEKPDILITFTPTVMSVUTEQ-----FNINIPVATVMTDYRLHKNWITPY 147
Db 1479 TREGLFVSELDPVLLIS-----KVMDDSKPSFGYISNNLKDVKVITKDNVN-----I 1526
QY 148 STRYVATKETQDFTDVGI-----DPSTVKVTGIPIDNKFETPINQKQWLNDNLDPPK 202
Db 1527 LTGYIL-----KDDIKISLSLTQDEKTIKLSNVHLD---ESGVAE----- 1564
QY 203 QTIKASAGFVSGKFDITMITDILAKSANAQVVMICGSKSKELKRSITAKFKLT----- 255
Db 1565 --ILKFMNRKGNNTSDSLMS--FLESMNIIKIFVNFLOSNIKIFLDANFIISGTSICQ 1620
QY 256 -----RWYLILGVTKHMNE--WMASSQLMITKP-----GGTIT- 287
Db 1621 FEFICDENNIQPYFIKFTNTLTNTLYVGNRQNMIVEPNYDLDDSGDTSVINFQSRY 1680
QY 288 -EGFARCIPMIFLNP-----APGOELENAF-----YPEKGFQKIADTPEEA 328
Db 1681 LGIDSCVKNVVISPNYIDEINIPVTNNTYPEVIVLDANYINEKINVNINDL---S 1737
QY 329 IKIVASLTNGNQLTNMISTMQDKI 354
Db 1738 IRYVWS-NDGNDFI--LMSTSEENKV 1760

RESULT 12

PCT-US92-00282-5

Sequence 5, Application PC/TUS9200282

GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.

APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-5

Query Match 4.5%; Score 93; DB 5; Length 531;
Best Local Similarity 21.2%; Pred. No. 0.58;
Matches 85; Conservative 64; Mismatches 147; Indels 104; Gaps 22;

QY 14 SFGNGHMQ-----VTOSIVNQLNDMLDHLVSIEHDLFMEAHPIILTSICKKWIINSFKYPR 69
Db 93 SFGNHFABERSFLTAPQTEYRNMMIVIGLYFINCOSLLQDRDTL-----NFFK---- 140
QY 70 NMYKGFYISRPDKLDKCF-----YKYVGLNKLINLIIKEKPDILILTF---PTP 115
Db 141 -----ESKFDALFDPALPCGVILAEYLGUPSV--YLFGRFPCSLHTFSKSPDP 188
QY 116 VMSV-----LPEQFNINIPVATVMTDYRLHKNWITPY-----STRYVATKETKQDFI 163
Db 189 VSYIPRCYTKFSDHMTFSQSVANFLV-----NLLEPYLYCYCLFSKYEKLASAVLKRD-V 241
QY 164 DVGIDPSTVKVVGIPIDNKFETP-----INQKWLIDNNLDPKDQOTILMSAGA 211
Db 242 DI-ITLSEVSWLLRYDFVLEYPRPVMNMFVIGGINCKK---RKDLSQEFAYINASGE 297
QY 212 FGV---SKGFDITMITDILAKSANAQVVMICGSKSKELKRSITAKFKLTMYLILGYTKHMN 268
Db 298 HGIVVFSLG--SNVSEIPEKKAMA--IADALGNQO---FVLWRYTGTFRSNLANNT--ILV 350
QY 269 EWMASSQLM-----ITKPGGITITEGFARCIPIFNIPAPGOELENAFYFEKGFQ- 319
Db 351 KWLPNQDLGLHPWTRAFITHAGSHGVYESICNGVPMVM-PLFGQMDMNAKRMETKGACV 409
QY 320 -----KTADTPEEAIKIVASLTNGNEQLTNMISTMQDK 353
Db 410 TLNVLEMTSEDLENALKAVINDKSYKENIMR--LSSLHKDR 448

RESULT 13

US-08-451-715A-4

```
; Sequence 4, Application US/08451715A
; Patent No. 5801013
; GENERAL INFORMATION:
; APPLICANT: Tao, Jianshi
; APPLICANT: Qui, Yan
; APPLICANT: Houman, Fariba
; APPLICANT: Shen, Xiaoyu
; APPLICANT: Schimmel, Paul R.
; TITLE OF INVENTION: Helicobacter Aminoacyl-tRNA Synthetase
; TITLE OF INVENTION: Proteins, Nucleic Acids and Strains Comprising Same
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,715A
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CP194-25
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 648 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-451-715A-4

Query Match 4.5%; Score 92; DB 1; Length 648;
Best Local Similarity 19.6%; Pred. No. 0.99;
Matches 80; Conservative 56; Mismatches 152; Indels 118; Gaps 19;

QY 4 QNKKLIITGSGFHGHQVQTQS--IVNQLNDMNDLHLSVIEHDLFMEHPILTICKWY 61
Db 38 QGEEVFELTGDEHG-QKIQSARLRNOSPAYADSIKFD-----QW- 81
QY 62 INSPKIFRMYKGFYISRPDKLDC-----FYK-----YYGLNK 95
Db 82 ---DFFNLVDGFTIRTDSEHQKVCQNAFIMEFKGDIYKGAISGYCYSCSYCAISK 137
QY 96 LIN-----LLKKEPDLILLFTPTVMSVLTEQFNINIPATVMTDRLH 140
Db 138 ADNTSDKVLCPDCLREALLSEESYFFKLSAYEKPLLEFYAKNPEAILPI-----YR-- 189
QY 141 KNWITPTSTRYVATKETQKQFIDVDGIDPSVKVTGPIDNKKFETPINOKO-WLIDNLD 199
Db 190 KNEVTSF-----IEQGLDLSITRISFE-WGIPLPKMNDDPKHVVYVWL----- 232
QY 200 PDKOTILMSAGAFVSGKGFDMITDILAKSANAQVVMICGK---SKELKRLSLAKFKLTR 256
Db 233 ---DALLNYASALGYLNGLD-----NKMAHECARHIVGKDLIRFHAIYWPFL 278
QY 257 MYLIGYTKHM--NEWMASSOLMTKGGTITGTFARCPIMFLNPAPGOELENAYFFE 314
Db 279 MSLNPLFLKQLCVHGWMTIEGVKMSKSLGNVLD---AQKLAMEY-----GLEELRYFLR 330
QY 315 EKGFGKIADTPPEAI--KIVASLNGNEQLNMLSTMEODKIKYATOT 360

; Sequence 4, Application US/08942012B
; Patent No. 6235278
; GENERAL INFORMATION:
; APPLICANT: Miller, Lois K.
; APPLICANT: Lu, Albert
; APPLICANT: Dierks, Peter
; APPLICANT: Black, Bruce
; TITLE OF INVENTION: Biological Insect Control Agents Expressing
; TITLE OF INVENTION: Insect-Specific Toxin Genes, Methods and Compositions
; FILE REFERENCE: 28-96a
; CURRENT APPLICATION NUMBER: US/08/942,012B
; CURRENT FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: 08/729,606
; PRIOR FILING DATE: 2000-10-01
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Spodoptera littoralis nuclear polyhedrosis virus
; US-08-942-012B-32

Query Match 4.4%; Score 91; DB 4; Length 515;
Best Local Similarity 24.0%; Pred. No. 0.87;
Matches 55; Conservative 33; Mismatches 101; Indels 40; Gaps 10;

QY 169 PSTVK-VTGIPIDNKFETPINQKWLIDNLDKQITILMSAGAFVSGKGFDMITDILA 227
Db 265 PSNVQYLGHIHPAVTSSVADE---IDNDLAEFLENSTM-GVVYVSLGSSVRSMDMS 318
QY 228 KSNA-----QVMICGSKSKELKRLSLAKFKLTRMYLILGYTKHNEWMASQLM 277
Db 319 NMLNVFETFRSIPYRVLWKVDSKIDFNIPSNVLQRFQRRVLKHRN-----VKVF 373
QY 278 ITRPGGITTEGFARCPIMFLNPAPGOELENAYFEKFGKIDRPEEAIVASLTN 337
Db 374 ITQGGVOSTDEADAGVPM-FGVPIMGDQFYNYMYTYGIGRGVDI-----LTV 422
QY 338 GNEQLTNMSTMEODKIKYATQTI-CRDLDLIGHSSQPOEIIYCKVPLY 385
Db 423 DARQLTEIVMDV-ADNEKYKNGTLWLRLDAI-----MDQPMRPLEKAVWY 465

; Sequence 4, Application US/09005180A
; Patent No. 612446
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN VPS35/MEM3/MEM3-RELATED PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,180A

; Sequence 4, Application US/09005180A
; Patent No. 612446
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN VPS35/MEM3/MEM3-RELATED PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,180A
```